OM of: US-09-303-518D-463 to: PIR_71:* out_format : pfs

Date: Jun 30, 2002 7:44 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142713_4317/app_query.fasta_1.23501
-DB=PIR_71 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=100 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09303518_@CGN1_1_1092
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-463
Query length: 1887
Database: PIR_71:*
Database sequences: 283138
Database length: 96089334
Search time (sec): 504.980000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
pir2:A81931	+	2200.00	2616.10	2.3e-138	556	probable adhesin NMA0853 [impor
pir2:B81176	+	1016.50	1207.07	9.3e-60	412	hypothetical protein NMB0647 [i
pir2:H81782	+	740.50	875.99	2.1e-41	498	adhesin MafB2 NMA2113 [imported
pir2:B82028	+	414.50	487.39	9.7e-20	482	adhesin NMA0324 [imported] - Ne
pir2:A43932	+	165.50	172.82	0.0051	3020	mucin 2 precursor, intestinal
pir2:G84693	+	163.50	182.09	0.0053	891	probable proline-rich protein [
pir2:T22808	+	162.00	181.69	0.0065	770	hypothetical protein F56H9.1 -
pir2:T45134	+	161.50	184.68	0.0064	529	hypothetical protein [imported]
pir2:T45025	+	158.00	162.27	0.0168	3570	mucin MUC5B, tracheobronchial
pir2:S57180	+	153.50	167.63	0.0260	1161	probable membrane protein YJR1
pir2:JQ0405	+	150.00	163.92	0.0440	1106	hypothetical 119.5K protein (u
pir2:T02345	+	150.00	159.32	0.0490	1791	hypothetical protein KIAA0324
pir2:JW0067	+	149.00	167.65	0.0457	660	chitinase (EC 3.2.1.14) A - Eme
pir2:T30944	+	148.00	156.50	0.0673	1873	surface protein precursor - En
pir2:S24169	+	147.00	173.05	0.0517	292	mucin - rat
pir2:A48018	+	147.00	170.61	0.0547	377	mucin 7 precursor, salivary - H
pir1:S48478	+	146.00	157.12	0.0851	1367	glucan 1,4-alpha-glucosidase (
pir2:I47141	+	145.50	165.61	0.0742	528	gastric mucin (clone PGM-2A) -
pir2:T14180	+	144.00	156.62	0.1106	1122	exiT protein - Mycobacterium s
pir2:B47328	+	143.50	153.21	0.1275	1507	natural killer cell tumor-recc
pir2:A53577	+	143.50	152.46	0.1298	1630	ascites sialoglycoprotein 1 -
pir2:T38459	+	143.50	151.55	0.1326	1794	hypothetical divergent repeat-
pir2:T29634	+	143.00	158.36	0.1203	825	hypothetical protein C12D12.1 -
pir2:T47182	+	142.00	152.84	0.1552	1299	hypothetical protein DKFZp434M
pir2:T30826	+	141.50	147.27	0.1883	2187	nascent polypeptide-associated
pir2:S35047	+	141.00	159.97	0.1487	543	mucin JUL7 - human
pir2:C39135	+	139.00	159.66	0.1923	437	hypothetical protein 3 (gyrB re
pir2:A46299	+	138.50	142.76	0.3045	2411	tyrosine kinase substrate disa
pir2:B41206	+	138.00	149.44	0.2771	1125	microtubule-associated protein
pir2:T15316	+	137.50	148.04	0.3047	1223	hypothetical protein B0302.1 -
pir2:S51342	+	137.00	151.30	0.3005	817	verprolin - yeast (Saccharomyce
pir1:VGBEX1	+	136.50	150.94	0.3226	797	glycoprotein X precursor - equi
pir2:T51023	+	136.00	138.88	0.4559	2649	hypothetical protein B7F21.40
pir2:E86185	+	135.00	155.68	0.3481	402	hypothetical protein [imported]
pir2:T15087	+	135.00	147.79	0.4190	919	hypothetical protein K06A5.4 -
pir2:T14342	+	135.00	137.91	0.5285	2588	NSD1 protein - mouse
pir2:T34369	+	134.50	140.90	0.5244	1777	hypothetical protein T19D12.1
pir2:I54632	+	134.00	142.74	0.5346	1377	tsh protein - Escherichia coli
pir2:T38236	+	134.00	141.24	0.5538	1611	hypothetical protein SPAC23A1.
pir2:T18535	+	133.50	143.85	0.5544	1151	high molecular mass nuclear an
pir2:H87520	+	133.00	152.79	0.4784	424	hypothetical protein CC2193
pir2:A29130	+	133.00	142.89	0.6036	1196	beta-amylase (EC 3.2.1.2) /
pir2:G01763	+	132.00	141.79	0.7018	1184	atrophin-1 - human
pir2:S50832	+	132.00	141.79	0.7018	1184	atrophin-1 - human
pir2:B84331	+	131.50	146.30	0.6721	694	hypothetical protein Vng1794
pir2:T43481	+	131.00	147.41	0.6969	580	probable mucin DKFZp434C196.
pir2:B81451	+	130.50	147.15	0.7464	560	flagellar M-ring protein Cj0
pir2:B46629	+	130.00	147.54	0.7873	505	mucin 6, gastric (3-repeat c
pir2:T48814	+	130.00	134.63	1.07	1952	hypothetical protein 15E6.2
pir2:A35938	+	129.50	132.69	1.19	2248	profilaggrin - human (fragm
pir2:E97631	+	129.00	146.91	0.9054	476	probable membrane transport
pir2:S29605	+	129.00	140.98	1.04	886	glycoprotein 350/220 - human
pir2:T14007	+	129.00	137.90	1.12	1224	microtubule-associated prot
pir1:W2WL8	+	128.50	145.88	0.9873	498	E2 protein - human papilloma
pir2:T42727	+	128.50	134.98	1.28	1560	proliferation potential-rel
pir2:T34248	+	128.00	140.79	1.18	798	hypothetical protein F31D5.4
pir2:T22696	+	128.00	140.17	1.20	851	hypothetical protein F55B11.
pir2:A53253	+	128.00	138.81	1.24	982	microtubule-associated prote
pir2:T31421	+	128.00	137.11	1.29	1173	C-terminal domain-binding p
pir2:S53363	+	127.50	150.22	1.01	279	mucin 5AC (clone JER58) - hu
pir2:I39540	+	127.50	143.88	1.17	542	chitinase (EC 3.2.1.14) II p
pir2:C81169	+	127.50	132.35	1.54	1815	IgA-specific metalloendopep
pir2:A53715	+	127.00	148.26	1.13	322	apomucin precursor - rat
pir2:S50755	+	127.00	144.59	1.23	473	hypothetical protein VSP-3 -
pir2:T33369	+	127.00	135.12	1.53	1275	hypothetical protein H02F09
pir2:I51618	+	126.50	136.94	1.56	990	nucleolar phosphoprotein - A
pir2:JC5630	+	126.50	134.19	1.67	1320	TCOF1 protein - mouse
pir2:S65763	+	126.00	143.37	1.43	474	chitinase (EC 3.2.1.14) prec
pir2:A81937	+	125.50	130.18	2.08	1773	IgA-specific metalloendopep
pir2:C86301	+	125.00	143.47	1.62	414	arginine/serine-rich protein
pir2:S35049	+	125.00	139.77	1.77	610	mucin JER57 - human (fragmen
pir2:S21323	+	125.00	138.74	1.81	680	probable endoglucanase - Rum
pir2:A41258	+	125.00	138.12	1.83	725	a-agglutinin core protein AG
pir2:T42731	+	125.00	135.00	1.97	1006	atrophin-1 related protein
pir2:B33135	+	125.00	131.34	2.15	1475	gtfB protein precursor - St
pir2:H84824	+	124.50	140.88	1.83	510	En/Spm-like transposon prote
pir2:T45462	+	124.50	135.83	2.06	866	membrane glycoprotein [impor
pir2:T21389	+	124.50	132.32	2.24	1251	hypothetical protein F26C11
pir2:E90696	+	124.50	130.84	2.32	1461	hypothetical protein ECs0054
pir2:A85547	+	124.50	130.84	2.32	1461	hypothetical protein Z0609
pir2:T31422	+	124.50	130.76	2.32	1473	C-terminal domain-binding p
pir2:S19150	+	124.00	138.30	2.07	628	hypothetical protein, 69K -
pir2:G84586	+	124.00	137.09	2.13	713	hypothetical protein At2g202
pir2:T21460	+	124.00	136.04	2.18	796	hypothetical protein ZK945.1
pir2:T30858	+	124.00	129.51	2.55	1577	glucosyltransferase - Strep
pir2:T43213	+	124.00	128.79	2.59	1701	ENBP1 protein - barrel medi
pir2:A34615	+	123.50	137.75	2.23	625	profilaggrin - rat (fragment
pir2:S07132	+	123.50	134.42	2.41	886	hypothetical protein 1 - Rho
pir1:QQBE21	+	123.50	134.20	2.43	907	membrane antigen gp350 - hum
pir2:I50114	+	123.00	139.08	2.30	511	early growth response protei
pir2:A44825	+	123.00	133.66	2.62	901	phosphoprotein, synapse-spec
pir2:T29776	+	123.00	132.24	2.71	1046	hypothetical protein C50F2.
pir2:C64221	+	123.00	132.18	2.71	1052	hypothetical 114K protein (
pir2:JT0345	+	123.00	129.63	2.88	1375	dextransucrase (EC 2.4.1.5)
pir2:JQ0110	+	122.50	136.51	2.60	628	hypothetical 69K protein - t
pir2:T21338	+	122.50	132.74	2.85	932	hypothetical protein F25D7.4
pir2:T19006	+	122.50	131.92	2.90	1016	ankyrin related protein C06
pir2:S61314	+	122.50	127.82	3.19	1561	IgA-specific metalloendopep
pir2:B45764	+	122.50	125.93	3.34	1902	lactocepin (EC 3.4.21.96) p
pir2:T49988	+	122.00	136.91	2.75	566	ovule development protein-li

seq_name: pir2:A81931

seq_documentation_block:
probable adhesin NMA0853 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: A81931
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491
A;Reference number: A81775; MUID:20222556

```
A;Accession: A81931
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-556 <PAR>
A;Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84134.1; PID:g737956
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: mafB3; NMA0853


alignment_scores:
            Quality: 2200.00              Length:    560
              Ratio:  4.593                Gaps:      8
 Percent Similarity:  85.536   Percent Identity: 77.500

alignment_block:
 US-09-303-518D-463 x A81931   ..

 Align seg 1/1  to: A81931  from: 1  to: 556

      1 TTGGGCATTTCCCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTG 50
        :::|||||||||||||||||||||||||||||||||||||||||||||||
      1 MetGlyIleSerArgLysIleSerLeuIleLeuSerIleLeuAlaValCy 17

     51 CCTGCCGATGCATGCACACGCCTCAGATTTGGCAAACGATCCCTTTATCC 100
        |||||||||||||||||||||||||||||||||||||||   |||||||
     17 sLeuProMetHisAlaHisAlaSerAspLeuAlaAsnAspSerPheIleA 34

    101 GGCAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTA 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     34 rgGlnValLeuAspArgGlnHisPheGluProAspGlyLysTyrHisLeu 50

    151 TTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGG 200
        |||||||||||||||||||||||   |||:::|||||||||||||||||||
     51 PheGlySerArgGlyGluLeuAlaGluArgSerGlyHisIleGlyLeuGl 67

    201 AAACATACAAAGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCG 250
        |||||||||||||||||||||||||:::|||:::|||||||||||||||:
     67 yAsnIleGlnSerHisGlnLeuGlyAsnLeuPheIleGlnGlnAlaAlaI 84

    251 TTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAA 300
        :::::|||||||||||||||||||||||||||||||||||||||||:::  .
     84 leLysGlyAsnIleGlyTyrIleValArgPheSerAspHisGlyHisGlu 100

    301 TTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGC 350
           ||||||||||||||||||||||||||||||||||||||||||||||||
    101 ValHisSerProPheAspAsnHisAlaSerHisSerAspSerAspGluAl 117

    351 CGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGAT 400
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    117 aGlySerProValAspGlyPheSerLeuTyrArgIleHisTrpAspGlyT 134

    401 ACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTAT 450
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    134 yrGluHisHisProAlaAspGlyTyrAspGlyProGlnGlyGlyGlyTyr 150

    451 CCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGT 500
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 ProAlaProLysGlyAlaArgAspIleTyrSerTyrAspIleLysGlyVa 167

    501 TGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAAC 550
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    167 lAlaGlnAsnIleArgLeuAsnLeuThrAspAsnArgSerThrGlyGlnA 184

    551 GGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTA 600
        |||||:::|||||||||||||||:::|||:::|||||||||||||||||
    184 rgLeuValAspArgPheHisAsnThrGlySerMetLeuThrGlnGlyVal 200

    601 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATC 650
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 GlyAspGlyPheLysArgAlaThrArgTyrSerProGluLeuAspArgSe 217

    651 GGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACA 700
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    217 rGlyAsnAlaAlaGluAlaPheAsnGlyThrAlaAspIleValLysAsnI 234

    701 TCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGT 750
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    234 leIleGlyAlaAlaGlyGluIleValGlyAlaGlyAspAlaValGlnGly 250

    751 ATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTC 800
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 IleSerGluGlySerAsnIleAlaValMetHisGlyLeuGlyLeuLeuSe 267

    801 CACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAAC 850
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    267 rThrGluAsnLysMetAlaArgIleAsnAspLeuAlaAspMetAlaGlnL 284

    851 TCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCC 900
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    284 euLysAspTyrAlaAlaAlaAlaIleArgAspTrpAlaValGlnAsnPro 300

    901 AATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCAT 950
        |||||||||||||||||||||||||||||||||||||   |||:::|||
    301 AsnAlaAlaGlnGlyIleGluAlaValSerAsnIlePheThrAlaValIl 317

    951 CCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCA 1000
        ||||:::|||||||||||||||||||||||||||||||||||||||||||
    317 eProValLysGlyIleGlyAlaValArgGlyLysTyrGlyLeuGlyGlyI 334

   1001 TCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCCG 1050
        ||||||||||||||||||||||||||||||   ||||||||||||
    334 leThrAlaHisProValLysArgSerGlnMetGlyGluIleAlaLeuPro 350

   1051 AAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAA 1100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    351 LysGlyLysSerAlaValSerAspAsnPheAlaAspAlaAlaTyrAlaLy 367

   1101 ATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGC 1150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    367 sTyrProSerProTyrHisSerArgAsnIleArgSerAsnLeuGluGlnA 384

   1151 GTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGC 1200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    384 rgTyrGlyLysGluAsnIleThrSerSerThrValProProSerAsnGly 400

   1201 AAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTT 1250
        ||||||||||||||||::::::|||||||||||||||   |||||||||
    401 LysAsnValLysLeuAlaAsnLysArgHisProLysThrLysValProPh 417

   1251 TGACGGTAAAGGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGA 1300
        ||||||||||||||||||||||||||||||   |||||||||||||||:
    417 eAspGlyLysGlyPheProAsnPheGluLysAspValLysTyrAspThrA 434

   1301 AGCTCGAT......ATTCAAGAATTATCGGGGGGCGGTATACCTAAGGCT 1344
        ::::::::      :::   :::::::::            |||
    434 rgIleAsnThrAlaValProGlnValAsn...........ProIleAsp 446

   1345 AAGCCTGTGTTTGATGCGAAA.................CCGAGATGGGA 1376
        :::||||||||||:::   |||                        ||||::
    447 GluProValPheAsnProLysGlySerValGlySerAlaHisSerTrpSe 463

   1377 GGTTGATAGGAAGCTTAAT......AAATTGACAACTCGTGAGCAGGTGG 1420
        ::::    :::::::::      ||||||      :::   :::::::
    463 rIleThrAlaArgIleGlnTyrAlaLysLeuProArgGlnGlyArgIleA 480

   1421 AG............AAAAATGTTCAGGAAACGAGA.............. 1443
        ::            ||||||   :::   :::
    480 rgTyrIleProProLysAsnTyrSerProSerAlaProLeuProLysGly 496

   1443 ..................................................... 1443

    497 ProAsnAsnGlyTyrLeuAspLysPheGlyAsnGluTrpThrLysGlyPr 513

   1444 .AGAAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAA.......... 1482
           |||:::::::::::::::|||:::      :::|||
```

```
 513 oSerArgThrLysGlyGlnGluPheGluTrpAspValGlnLeuSerLysT 530

1483 .....CGAGAA.........TGGGAAAATAAAACAGGGTTAGATTTTAAT 1518
          ||||||         |||   ::::::   |||      :::|||
 530 hrGlyArgGluGlnLeuGlyTrpAlaSerArgAspGlyLysHisLeuAsn 546

1519 CATTTTATAGGTGGTGATATCAATAAGAAA 1548
           |||   |||   |||:::   |||
 547 IleSerIleAspGlyLysIleThrHisLys 556

seq_name: pir2:B81176

seq_documentation_block:
hypothetical protein NMB0647 [imported] - Neisseria meningitidis (strain MC58 serogroup
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: B81176
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755
A;Accession: B81176
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-412 <TET>
A;Cross-references: GB:AE002419; GB:AE002098; NID:g7225863; PIDN:AAF41068.1; PID:g722587
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB0647
C;Superfamily: Neisseria meningitidis hypothetical protein NMB0647


alignment_scores:
           Quality: 1016.50              Length:    449
             Ratio:  3.592                Gaps:     12
 Percent Similarity:  63.029   Percent Identity: 51.448

alignment_block:
 US-09-303-518D-463 x B81176   ..

 Align seg 1/1  to: B81176  from: 1  to: 412

 588 GACGCAAGGAGTAGGCGACGGATT..................CAAACGCG 619
     |||   ::::::   |||:::|||                  |||:::::
  42 AspLysGlnThrAlaArgLysIleTyrSerValIleLysGluGlnValLy 58

 620 CCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCC 669
     :|||      :::      :::|||:::|||   :::||||||:::::::
  58 sHisGluCysSer...SerThrGlyArgIleThrGluCysArgGlnAsnI 74

 670 TTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGA 719
     ::   |||       ::::::   |||
  74 leGlyArgIleIleGluPheThrGlnAsp..................... 83

 720 AATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACA 769
               :::|||      :::   :::|||         |||::
  84 ............LysArgPheAspSerArgPheLysAspLeuLysLysGl 96

 770 TTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCG 819
     :                    |||   ::::::    ||||||
  96 u.....................SerLeuTyrTyrLeu.AsnLysHisPro 105

 820 CGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAGC 869
           ||||||:::                    :::|||
 106 .........AspLeuValAlaSerTyrLeuLysAlaGluTyrGluLysLe 119

 870 AGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAG 919
           :::|||   ::::::   :::         :::   |||   |
 119 uAspArgGluAspLysSerIleLeuHisArgTyrIleSerProGlyAlaG 136

 920 AAGCCGTCAGC...AATATCTTTATGGCAGCCATCCCCATCAAAGGGATT 966
     ||   ||||||   ::::::   ::::::         :::   |||
 136 luIleValSerGlySerLeuGlyValValLeuSerGlyValAlaGlyGly 152

 967 GGAGCTGTCCGGGGAAAATACGGCTTGGGC.................GG 998
     |||:::            :::|||||||||                ||
 153 GlySerCysAlaGluThrPheGlyLeuGlyCysAlaAlaAlaLeuValGl 169

 999 CATCACGGCA.........CATCCTGTCAAGCGGTCGCAGATGGGCGCGA 1039
     |:::|||:::         |||   :::      ::::::
 169 yValThrSerSerTyrAspHisValIleThrGlyThrLys.......... 182

1040 TCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCG 1089
                                          |||||||:::
 183 ................................AsnPheGlyLysLys 187

1090 GCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAA 1139
     |||       :::   |||
 188 AlaSerGluGlnArgPro.............................. 193

1140 CTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGC 1189
                                              |||:::
 194 ....................................ThrIleAlaV 197

1190 CGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACA 1239
       :::      |||::::::   |||:::::|||
 197 alGlnAlaLeuLysGlnLeuGlyLeuSerGluGlnAla............ 209

1240 GGCGTACCGTTTGACGGTAAAGGGTTTCCGAATTTTGAGAAGCACGTGAA 1289
                                              ::::::|||::
 210 .................................AlaGluTyrValGl 214

1290 ATATGATACGAAGCTCGATATTCAAGAATTATCGGGGGGCGGTATACCTA 1339
     :::::::      |||                |||||||||||||||||||
 214 nPheSerIleAspLeuPheSerValGlyLysSerGlyGlyGlyIleProL 231

1340 AGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAGGTTGATAGGAAG 1389
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 231 ysAlaLysProValPheAspAlaLysProArgTrpGluValAspArgLys 247

1390 CTTAATAAATTGACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGAAAC 1439
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 248 LeuAsnLysLeuThrThrArgGluGlnValGluLysAsnValGlnGluTh 264

1440 GAGAAGAAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAGAAT 1489
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 264 rArgArgArgSerGlnSerSerGlnPheLysAlaHisAlaGlnArgGluT 281

1490 GGGAAAATAAAACAGGGTTAGATTTTAATCATTTTATAGGTGGTGATATC 1539
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 281 rpGluAsnLysThrGlyLeuAspPheAsnHisPheIleGlyGlyAspIle 297

1540 AATAAGAAAGGCACAGTAACAGGAGGGCATAGTCTAACCCGTGGTGATGT 1589
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 298 AsnLysLysGlyThrValThrGlyGlyHisSerLeuThrArgGlyAspVa 314

1590 ACGGGTGATACAACAAACCTCGGCACCTGATAAACATGGGGTTTATCAAG 1639
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 314 lArgValIleGlnGlnThrSerAlaProAspLysHisGlyValTyrGlnA 331

1640 CGACAGTGGAAATTAAAAAGCCTGATGGAAGTTGGGAGGTGAAAACGAAA 1689
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 331 laThrValGluIleLysLysProAspGlySerTrpGluValLysThrLys 347

1690 AAAGGTGGGAAAGTGATGACCAAGCACACCATGTTCCCAAAAGATTGGGA 1739
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 348 LysGlyGlyLysValMetThrLysHisThrMetProLysAspTrpAs 364

1740 TGAGGCTAGAATTAGGGCTGAAGTTACTTCGGCTTGGGAAAGTAGAATAA 1789
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 364 pGluAlaArgIleArgAlaGluValThrSerAlaTrpGluSerArgIleM 381

1790 TGCTTAAGGATAATAAATGGCAGGGTACAAGTAAATCGGGTATTAAAATA 1839
```

```
     |||||||||||||||||||||||||||||||||||||||||||||||||
  381 etLeuLysAspAsnLysTrpGlnGlyThrSerLysSerGlyIleLysIle 397

 1840 GAAGGATTTACCGAACCTAATAGAACAGCATATCCCATTTATGAA 1884
      ||||||||||||||||||||||||||||||||||||||||||||
  398 GluGlyPheThrGluProAsnArgThrAlaTyrProIleTyrGlu 412
```

seq_name: pir2:H81782

seq_documentation_block:
adhesin MafB2 NMA2113 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: H81782
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556
A;Accession: H81782
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-498 <PAR>
A;Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85327.1; PID:g738073
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: mafB2; NMA2113

alignment_scores:
Quality: 740.50 Length: 586
Ratio: 2.165 Gaps: 18
Percent Similarity: 58.362 Percent Identity: 32.253

alignment_block:
US-09-303-518D-463 x H81782 ..

Align seg 1/1 to: H81782 from: 1 to: 498

```
   28 ATTCTGTCCATACTGGCAGTGTGC.................CTGCCGAT 59
      :::::::::::::|||||||:::|||                |||   ::
    7 LeuIleLysLeuLeuAlaAlaCysAlaValAlaAlaAlaAlaLeuIleGl 23

   60 GCATGCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTC 109
      :   |||   |||:::||||||||||:::|||||||||||   :::
   23 nProAlaLeuAlaAlaAspLeuAlaGlnAspProPheIleThrAspAsnA 40

  110 TCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGC 159
        :::|||||||||:::||||||   ||||||||||||||||||||:::
   40 laGlnArgGlnHisTyrGluProGlyGlyLysTyrHisLeuPheGlyAsp 56

  160 ...AGGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACAT 206
         ||||||:::::::::   |||:::|||:::|||::::::   :::
   57 ProArgGlySerValSerAspArgThrGlyGlnIleAsnValIleGlnAs 73

  207 ACAAAGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAG 256
        :::|||:::::::|||:::|||:::||||||||||||   :::::::|
   73 pTyrThrHisArgMetGlyAsnLeuLeuIleGlnGlnAlaAsnIleAsnG 90

  257 GAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCAT 306
      ||:::|||||||||   :::|||||||||   ||||||::::::   |||
   90 lyThrIleGlyTyrHisThrArgPheSerGlyHisGlyTyrGluGluHis 106

  307 TCGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAG 356
      :::||||||||||||||||||::   |||   |||:::|||   |||::
  107 AlaProPheAspAsnHisAlaAlaAspSerAlaSerGluGluLysGlyAs 123

  357 TCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAAC 406
      :      :::||||||:::::::||||||:::::::|||:::|||:::||||
  123 nValAspGluGlyPheThrValTyrArgLeuAsnTrpGluGlyHisGluH 140

  407 ACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCT 456
      ||||||||||||||:::||||||||||||:::|||||||:::|||||||
```

```
  140 isHisProAlaAspAlaTyrAspGlyProLysGlyGlyAsnTyrProLys 156

  457 CCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCA 506
      |||   ||||||||||||   |||:::|||   ::::::|||:::|||::
  157 ProThrGlyAlaArgAspGluTyrThrTyrHisValAsnGlyThrAlaAr 173

  507 AAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCTTG 556
      ::::|||:::|||||||   ||||||:::|||||||      ||||||::::
  173 gSerIleLysLeuAsnProThrAspThrArgSerIleArgGlnArgIleS 190

  557 CCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGAC 606
      ::|||::::::   |||   |||:::   :::::::::   :::   :::
  190 erAspAsnTyrSerAsnLeuGlySerAsnPheSerAspArgAlaAspGlu 206

  607 GGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAA 656
      :::   ::::::      :::::::::   ::::||||||||||   |||||
  207 AlaAsnArgLysMetPheGluHisAsnAlaLysLeuAspArgTrpGlyAs 223

  657 TGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATCG 706
      |:::   |||   :::|||||||:::|||      :::   |||   :::
  223 nSerMetGluPheIleAsnGlyValAlaAlaGlyAlaLeuAsnProPheI 240

  707 GCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGC 756
        :::|||||||||   :::|||   ||||||   :::   |||
  240 leSerAlaGlyGluAlaLeuGlyIleGlyAspIleLeuTyrGlyThrArg 256

  757 GAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGA 806
      :::   :::   |||:::|||:::::::::::   |||   :::||
  257 TyrAlaIleAspLysAlaAlaMetArgAsnIleAlaProLeuProAlaGl 273

  807 AAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAG 856
      |:::|||:::|||   |||:::   |||:::::::::|||   ::::::
  273 uGlyLysPheAlaValIleGlyGlyLeuGlySerValAlaGlyPheGluL 290

  857 ACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCC 906
        :::      |||:::      |||      :::||||||||||||
  290 ysAsnThrArgGluAlaValAspArgTrpIleGlnGluAsnProAsnAla 306

  907 GCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCAT 956
      |||:::   :::|||||||:::   |||:::::      |||
  307 AlaGluThrValGluAlaLeuValAsnValLeuProPheAla........ 320

  957 CAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCATCACGG 1006

  320 ..................................................... 320

 1007 CACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGG 1056
                |||   :::::::::          |||   |||   |||
  321 ..........LysValLysAsnLeuThrLysAlaAlaLysPro...Gly 332

 1057 AAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAAATACCC 1106
      |||:::|||||||||   :::|||:::   |||
  333 LysAlaAlaValSerGlyAspPheSerAlaAla.................. 343

 1107 GTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACG 1156
             |||:::::::|||:::   |||:::        :::        |
  344 .......TyrAsnThrArgThrThrArgLysValThrThrGluThrGluG 358

 1157 GCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAAT 1206
      ||   ::::::|||                        :::|||   ||||||
  358 lyLeuAsnArgIleArg.................GlnAsnGlnLysAsn 368

 1207 GTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGG 1256
      ::::::   ::::::::::::                            ||
  369 SerAsnIleHisGluLysAsnTyr.......................Gl 377

 1257 TAAAGGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCG 1306
      |:::      ||||||         |||::::::      :::      :
  377 yArgAspAsnProAsn.........HisIleAsnValLeuSerGlyAsnS 391

 1307 ATATTCAAGAATTA..............TCGGGGGGCGGTATACCTAAG 1341
      ::||||||:::::                :::||||||||||
  391 erIleGlnHisIleLeuTyrGlyAspGluAlaGlyGlyGly......... 404
```

```
 1342 GCTAAGCCTGTGTTTGATGCGAAACCG.....................AG 1370
               :::|||   :::||||||                      ::
  405 ......HisLeuPheProGlyLysProGlyLysThrThrPheProGlnHi 419

 1371 ATGGGAGGTTGATAGGAAGCTTAATAAATTGACAACTCGTGAGCAGGTGG 1420
      :|||::::::           :::|||:::|||
  419 sTrpSerAla............SerLysIleThr................ 426

 1421 AGAAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGAGTAGTCAGTTTAAA 1470
                                    :::   |||::::::
  427 ............................HisGluIleSerAspIleVal 433

 1471 GCCCATGCGCAACGAGAATGGGAAAATAAAACAGGGTTAGATTTTAATCA 1520
      :::      :::   :::|||      :::||||||
  434 ThrSerProLysThrGlnTrpTyrAlaGlnThrGly.............. 445

 1521 TTTTATAGGTGGTGATATCAATAAGAAAGGCACAGTAACAGGAGGGCATA 1570
                                        |||||||||
  446 ....................................ThrGlyGlyLysT 450

 1571 GTCTAACCCGTGGTGATGTACGGGTGATACAACAAACCTCGGCACCTGAT 1620
        :::::::::|||      |||      :::   :::|||
  450 yrIleAlaLysGly......ArgProAlaArgTrpValSerTyrGluThr 464

 1621 AAACATGGGGTTTATCAAGCGACAGTGGAAATTAAAAAGCCTGATGGAAG 1670
      :::   |||:::         ||||||
  465 ArgAspGlyIleArgIleArgThrVal...................... 473

 1671 TTGGGAGGTGAAAACGAAAAAAGGTGGGAAAGTGATGACCAAGCACACCA 1720
       :::|||      |||         |||||||||:::         |||
  474 .TyrGluProAlaThr.........GlyLysValVal.........ThrA 484

 1721 TGTTCCCA 1728
        ||||||
  484 laPhePro 486
```

```
seq_name: pir2:B82028

seq_documentation_block:
adhesin NMA0324 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: B82028
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556
A;Accession: B82028
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-482 <PAR>
A;Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83629.1; PID:g737908
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: mafB; NMA0324


alignment_scores:
            Quality:  414.50              Length:    457
              Ratio:  1.692                Gaps:     20
 Percent Similarity:  53.611   Percent Identity: 30.416

alignment_block:
 US-09-303-518D-463 x B82028   ..

 Align seg 1/1  to: B82028  from: 1  to: 482

      31 CTGTCCATACTGGCAGTGTGCCTGCCGATGCATGCACACGCCTCAGATTT 80
         :::|||:::|||   :::   :::          :::|||||||:::   ||
      16 IleSerLeuLeuGlnIleProIle.........SerHisAlaAsnGlyLe 29

      81 GGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAAC 130
         |      |||      :::|||:::   ::::::   :::|||:::||||
      29 u......AspAlaArgLeuArgAspAspMetGlnAlaLysHisTyrGluP 44

     131 CCGACGGGAAATACCACCTATTCGGC...AGCAGGGGGGAGCTTGCCNAG 177
         ||   ||||||||||||||||||||   :::||||||:::::
      44 roGlyGlyLysTyrHisLeuPheGlyAsnAlaArgGlySerVal...... 58

     178 CGCAACGGCCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCA 227
         :::|||          ::::::   ::::::::::::      :::|||
      59 LysAsnArgValTyrAlaValGlnThrPheAspAlaThrAlaValGlyPr 75

     228 CCTGATG.........ATTCAACAGGCGGCCGTTGAAGGAAATATCGGCT 268
           ::::::           ::::::::::::   ||||||   |||||||
      75 oIleLeuProIleThrHisGluArgThrGlyPheGluGlyIleIleGlyT 92

     269 ACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGAC 318
         ||   ::::::||||||   ||||||||:::   ||||||||||||||||
      92 yrGluThrHisPheSerGlyHisGlyHisGluValHisSerProPheAsp 108

     319 AACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGG 368
         ||||||   |||   |||   ||||||   :::|||:::          ||
     109 AsnHisAspSerLysSerThrSerAspPheSerGlyGlyValAspGlyGl 125

     369 ATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCG 418
         ||||::::::|||::::::|||      |||   |||   ||||||   |
     125 yPheThrValTyrGlnLeuHisArgThrGlySerGluIleHisProGluA 142

     419 ACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCG 468
         ||||||||||||||||||||||:::   ||||||   |||   ||||||
     142 spGlyTyrAspGlyProGlnGlySerAspTyrProProProGlyGlyAla 158

     469 AGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCT 518
         ||||||||||||||||||   :::||||||::::::   :::   :::
     159 ArgAspIleTyrSerTyrTyrValLysGlyThrSerThrLysThrLysSe 175

     519 CAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTC. 567
          |||::::::   :::                 ::::::||||||:::
     175 rAsnIleValProArgAlaPro...........PheSerAspArgTrpL 188

     568 .....CACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTC 612
             :::||||||||||||                 ::::::||||||
     188 euLysGluAsnAlaGlyAla.................AlaSerGlyPhe 198

     613 .......................AAACGCGCCACCCGATACAGCCCCGA 638
                                 |||         :::   :::|||::
     199 PheSerArgAlaAspGluAlaGlyLysLeuIleTrpGluSerAspProAs 215

     639 GCTGGACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATA 688
         :   :::      :::|||      :::   ::::::|||   ::::::
     215 nLysAsnTrpTrpAlaAsnArgMetAspAspIleArgGlyIleValGlnG 232

     689 TCGTCAAAAAC......ATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 732
           :::   |||      ::::::|||      |||      :::|||||||
     232 lyAlaValAsnProPheLeuMetGlyPheGlnGlyValGlyIleGlyAla 248

     733 ......GGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGT 776
               :::|||||||:::   :::::::::   :::         ::
     249 IleThrAspSerAlaValSerProValThrAspThrAlaAlaGlnGlnTh 265

     777 CATGCACGGC.........TTGGGTCTGCTTTCCACCGAAAACAAGATGG 817
         :::::::|||         ||||||   ||||||   |||   ::::::|
     265 rLeuGlnGlyIleAsnHisLeuGlyAsnLeuSerProGluAlaGlnLeuA 282

     818 CGCGCATCAACGATTTGGCAGATATGGCG...CAACTCAAAGACTATGCC 864
         ||       :::   |||   |||   |||       :::||||||
     282 laAlaAlaThrAlaLeuGlnAspSerAlaPheAlaValLysAspGlyIle 298

     865 GCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGG 914
           :::|||   |||:::||||||      :::||||||   :::
     299 AsnSerAla...ArgGlnTrpAlaAspAlaHisProAsnIleThrAlaTh 314

     915 CATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGG. 963
```

```
          :::::::::      :::      ||||||      :::   |||
    314 rAlaGlnThrAlaLeuAlaValAlaGluAlaAlaThrThrValTrpGlyG 331

    963 ..................................................  963

    331 lyLysLysValGluLeuAsnProThrLysTrpAspTrpValLysAsnThr 347

    964 .........ATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCATCAC 1004
                 |||||||||                              ||
    348 GlyTyrLysThrProAlaValArg.......................Th 356

   1005 GGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCCGAAAG 1054
        |   |||   :::      ::::::   |||:::         |||.
    356 rMetHisThrLeuAspGlyGluMetAlaGlyGlyAsnArgProPro.... 371

   1055 GGAAATCCGCCGTCAGCGACAATTTTGCCGATGCG.............. 1089
          ||||||   :::|||::::::   |||||||||
    372 ..LysSerIleThrSerAsnSerLysAlaAspAlaSerThrGlnProSer 387

   1090 ....................................GCATACGC 1097
                                           ||||||
    388 LeuGlnAlaGlnLeuIleGlyGluGlnIleSerSerGlyHisAlaTyrAs 404

   1098 CAAATAC.........CCGTCCCCTTACCATTCCCGAAATATCCGTTCAA 1138
         |||:::              :::   :::   :::   ||||||:::|||
    404 nLysHisValIleArgGlnGlnGluPheThrAspLeuAsnIleAsnSerP 421

   1139 ACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCG 1188
             :::               |||||||||:::|||
    421 roAlaAspPheAlaArgHisIleGluAsnIleValSerHis......... 434

   1189 CCGTCAAACGGCAAAAATGTC 1209
        |||:::|||   |||::::::
    435 ProThrAsnMetLysGluLeu 441
```

seq_name: pir2:A43932

seq_documentation_block:
mucin 2 precursor, intestinal - human (fragments)
N;Alternate names: mucin SMUC-41
C;Species: Homo sapiens (man)
C;Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C;Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A;Reference number: A49963; MUID:94132002
A;Accession: A49963
A;Molecule type: mRNA
A;Residues: 1-639 <GU1>
A;Cross-references: GB:L21998
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
A;Reference number: A45106; MUID:93016075
A;Accession: A45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 626-1895 <GU2>
A;Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A;Note: sequence extracted from NCBI backbone (NCBIP:116706)
A;Accession: B45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 2037-3020 <GU3>
A;Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
A;Experimental source: colon
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)
R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp
A;Reference number: A43932; MUID:91358717
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350,'L',1352-1411,'S',1413-1448,'P',1450-1503,'T',1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
A;Reference number: A33532; MUID:89197956
A;Accession: B33532
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
J. Clin. Invest. 87, 77-82, 1991
A;Title: Human bronchus and intestine express the same mucin gene.
A;Reference number: A61257; MUID:91086481
A;Accession: A61257
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>
A;Experimental source: bronchus
R;Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstn
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A;Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the
A;Reference number: PQ0328; MUID:92198477
A;Accession: PQ0328
A;Molecule type: mRNA
A;Residues: 2328-2468 <XUG>
A;Cross-references: GB:M86523
A;Experimental source: small intestine
A;Accession: PQ0329
A;Molecule type: protein
A;Residues: 2328-2342,'K',2344-2354 <XUG1>
C;Genetics:
A;Gene: GDB:MUC2
A;Cross-references: GDB:120203; OMIM:158370
A;Map position: 11p15.5-11p15.5
C;Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; v
C;Keywords: glycoprotein; intestine; tandem repeat
F;2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>

alignment_scores:
            Quality: 165.50   Length: 334
              Ratio: 1.226   Gaps: 11
 Percent Similarity: 40.419   Percent Identity: 22.754

alignment_block:
 US-09-303-518D-463 x A43932   ..

 Align seg 1/1 to: A43932 from: 1 to: 3020

```
     269 ACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGAC 318
         |||   ::      |||   |||   |||::::::      |||   ||
    1520 ThrThrThrProSerProProThrThrThrThrThrProProProTh 1536

     319 AACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGG 368
         ||||   |||      |||      |||   |||::::::|||   |||:
    1536 rThrThrProSer...ProProThrThrThrProIleThrProProThrS 1552

     369 ATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCG 418
         :::::      |||:::         |||   |||   :::   ||||||
    1552 erThr......ThrThrLeuProProThrThrThrProSerProProPro 1566

     419 ACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCG 468
         |||:::   |||            ::::::   |||   |||
    1567 ThrThrThrThrThrProProProThrThrThrProSerPro........ 1580

     469 AGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCT 518
                       :::||||||      :::   |||   |||::::::::
    1581 .........ProThrThrThrThrProSerProProThrIleThrThrT 1594
```

```
 519 CAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCC 568
     ::|||   |||   |||::::::|||:::          |||::::::
1594 hrThrProProProThrThrThrProSerProProThrThrThrThrThr 1610

 569 ACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGC 618
     |||   |||                          |||   :::   ::
1611 ThrProPro.....................ProThrThrThrProSe 1619

 619 GCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGC 668
     :|||||   |||:::|||              :::           |
1619 rProProThrThrThrProIleThrProProThrSerThrThrThrLeuP 1636

 669 CTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAG 718
     ||   |||:::                  |||::::::::::
1636 roProThrThrThrProSerProProProThrThrThrThrThrProPro 1652

 719 AAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAAC 768
         ::::::   :::   |||      :::      :::
1653 ProThrThrThrProSerProProThrThrThrThrProSerProProIl 1669

 769 ATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGC 818
         :::   |||              ||||||   |||
1669 eThrThrThrThrThr.............ProProProThrThrThrP 1681

 819 GCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAG 868
       :::|||   |||                      |||   |||:::
1681 roSerSerProIle...................ThrThrThrProSer 1690

 869 CAGCCA.....................................TC 876
       |||                                     ||
1691 ProProThrThrThrMetThrThrProSerProThrThrThrProSerSe 1707

 877 CGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGT 926
     |   |||      :::   ||||||          :::   :::|||
1707 rProIleThrThrThrThrThrProSerSerThrThrThrProSerProP 1724

 927 CAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCC 976
         ::::::      |||||||||:::
1724 roProThrThrMetThrThrProSerProThr.................. 1734

 977 GGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCG 1026

1734 ..................................................... 1734

1027 CAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAA 1076
                          :::||||||:::::|||:
1735 ....................ThrThrProSerProProThrThrThrM 1744

1077 TTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATT...... 1120
     ::      :::      |||   ::::::   |||||||||
1744 etThrThrLeuProProThrThrThrSerSerProLeuThrThrThrPro 1760

1121 ......CCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAA 1164
           |||:::|||:::      |||:::|||      ::::::
1761 LeuProProSerIleThrProProThrPheSerProPheSerThr..... 1775

1165 AACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACT 1214
      |||:::|||      ||||||           :::   :::   ||||
1776 .ThrThrProThrThrProCys...........ValProLeuCysAsnT 1788

1215 GG 1216
     ||
1788 rp 1788
```

seq_name: pir2:G84693

seq_documentation_block:
probable proline-rich protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: G84693
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: G84693
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-891 <STO>
A;Cross-references: GB:AE002093; NID:g3980411; PIDN:AAC95214.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g29210
A;Map position: 2

alignment_scores:
Quality: 163.50 Length: 599
Ratio: 0.654 Gaps: 30
Percent Similarity: 41.736 Percent Identity: 23.706

alignment_block:
US-09-303-518D-463 x G84693 ..

Align seg 1/1 to: G84693 from: 1 to: 891

```
 255 AGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCC 304
     |||:::   |||:::|||   |||   |||||||||||:::::   ||
 286 ArgArgArgArgIleHisSerPro...PheArgSerArgSerArgSerPr 301

 305 ATTCGCCCTT...CGACAACCATGCCTCACATTCCGATTCTGACGAAGCC 351
     |:::         |||:::|||            |||
 301 oIleArgArgHisArgArgProThrHisGluGlyArgArgGlnSerProA 318

 352 GGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATA 401
           ||||||   |||   :::|||   |||
 318 laProSerArgArgArgArgSerProSerPro..... 328

 402 CGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATC 451
           |||:::|||||||||                  |||      ||
 329 ......ProAlaArgArgArg.................ArgSerProSe 337

 452 CCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTT 501
     |         ||||||                           ||||||
 337 rProProAlaArgArg.........................ArgArgS 345

 502 GCCCAAAATATCCGCCTCAACCTGACCGACAACCG.............. 536
       |||:::   |||   :::      |||:::|||
 345 erProSerProProAlaArgArgHisArgSerProThrProProAlaArg 361

 537 CAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGC 586
     |||:::|||:::              |||   |||   :::|||||||:::
 362 GlnArgArgSer............ProSerProProAlaArgArgHisAr 374

 587 TGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCC 636
      :::             ||||||         |||   |||      |||
 374 gSerProProProAlaArgArg......ArgArgSerProSerProPro. 388

 637 GAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGA 686
                   ::::::   |||||||||   :::      :::|||
 389 .............AlaArgArgArgArgSerProSerProProAlaArg 400

 687 TATCGTCAAAAACATCATCGGCGCGGCAGGAGAAAT.............. 722
        |||::::::              |||||||||
 401 ArgArgArgSerProSerProLeuTyrArgArgAsnArgSerProSerPr 417

 723 ....TGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAAC 768
           ||||||:::|||   |||:::         :::|||   :::
 417 oLeuTyrArgArgAsnArgSerArgSerProLeuAlaLysArgGlyArgS 434

 769 ATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGAT... 815
               :::|||      |||         |||   :::|||
 434 erAspSerProGlyArgSerProSerProValAlaArgLeuArgAspPro 450
```

```
 816 ...GGCGCG.....................CATCAACGATTTGGCAGATA 841
       ||||||                      :::|||||||:::
 451 ThrGlyAlaArgLeuProSerProSerIleGluGlnArgLeuProSerPr 467

 842 TGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTC 891
       :::::|||||||||                  |||   |||    |
 467 oProValAlaGlnArgLeuProSerProProProArgArgAlaGlyLeuP 484

 892 CAAAACCCCAATGCCGCACAAGGCATAGAAGC................... 923
     ||:::|||           |||      |||
 484 roSerProProProAlaGlnArgLeuProSerProProProArgArgAla 500

 924 .........CGTCAGCAATATCTTTATGGCAGCCAT.............. 950
                    :::   :::   |||||||||
 501 GlyLeuProSerProMetArgIleGlyGlySerHisAlaAlaAsnHisLe 517

 951 .......CCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTT. 992
           |||         :::   |||   ||||||:::   ::::::
 517 uGluSerProSerProSerSerLeuSerProProGlyArgLysLysValL 534

 992 ................................................. 992

 534 euProSerProProValArgArgArgArgSerLeuThrProAspGluGlu 550

 993 ....................GGGCGGCATCACGGCACATCCTGTCAAGC 1021
                         |||||||||  |||:::   |||
 551 ArgValSerLeuSerGlnGlyGlyArgHis.ThrSer...ProSerHisI 566

1022 GGTCGCAGATGGGCGCGATCGCA......TTGCCGAAAGGGAAATCCGCC 1065
       :::|||   |||:::::::::            :::||||||||:::
 566 leLysGlnAspGlySerMetSerProValArgGlyArgGlyLysSerSer 582

1066 GTCAGCGACAATTTTGCCGATGCG.........GCATACGCCAAATACCC 1106
       |||::::::         |||      ::      ::   ||
 583 ProSerSerArgHisGlnLysAlaArgSerProValArgArgArgSerPr 599

1107 GTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT.... 1152
     |:::|||   :::   |||:::   ||||||:::      :::|||
 599 oThrProValAsnArgArgSerArgArgSerSerSerAlaSerArgSerP 616

1153 .....TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAAC 1197
               :::::::::      ||||||:::      ||||||:::
 616 roAspArgArgArgArgArgSerProSerSerSerArgSerProSerArg 632

1198 GGCAAAAATGTCAAACTG................................ 1215
     :::::::::     :::
 633 SerArgSerProProValLeuHisArgSerProSerProArgGlyArgLy 649

1216 .......GCAGACCAACGCCACCCG....................... 1233
              :::::::|||   |||
 649 sHisGlnArgGluArgArgSerProGlyArgLeuSerGluGluGlnAspA 666

1234 .........................AAGACAGGCGTACCGTTTGACGGT 1257
                              :::|||:::|||
 666 rgValGlnAsnSerLysLeuLeuLysArgThrSerVal............ 678

1258 AAAGGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGA 1307
               |||:::   :::|||:::   :::         ||||||
 679 .........ProAspThrAspLysArgLysGlnLeuProGluLysLeu.. 691

1308 TATTCAAGAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTG 1357
               |||:::   |||   :::         |||
 692 ..........LeuGluValGlyArgValGluHisTyrLys.......... 701

1358 ATGCGAAACCGAGATGGGAGGTTGATAGGAAGCTTAATAAATTGACAACT 1407
                     |||   :::|||||||   :::||||||:::
 702 .................GluGlnGluArgLysSerAspLysLeuSerGlu 712

1408 CGTGAGCAGGTGGAGAAAAAT.........................GT 1430
     :::::::::|||:::::::::
 713 LysArgSerValHisArgHisHisGlySerGlnMetSerProValGluAs 729

1431 TCAGGAAACGAGAAGAAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCGC 1480
```

```
          :::|||   |||   |||      ||||||:::   |||         |
  729 nSerGluGlyArgSerArgProValSerSerLysValLysAspSerGluG 746

 1481 AACGAGAATGGGAAAATAAAACAGGGTTAGATTTTAATCATTTTATAGGT 1530
       ||   |||   |||:::::::::   ||||||   |||
  746 lnValGluLysGluAspAsnSerAspLeuAspAlaAsnLeu......... 759

 1531 GGTGATATCAATAAGAAAGGCACAGTAACAGGAGGGCATAGTCTA..... 1575
         :::   ::::::|||   |||:::         |||::::::
  760 ...SerCysAspSerLysAspThrIleArg......HisGlnIleLysAs 773

 1576 .......ACCCGTGGTGATGTACGGGTGATACAACAAACCTCGGCACCTG 1618
                 |||   :::   |||      ::::::      :::   |
  773 pLysAsnArgArgLysAsnLysArgSerSerArgGluGluValSerSerA 790

 1619 ATAAACATGGGGTTTATCAAGCGACAGTGGAAATTAAAAAGCCTGATGGA 1668
       ||   :::|||      ::::::   |||:::   :::|||
  790 spAspAsnGlySerSerAspSerAspValAspAspArgLys......... 803

 1669 AGTTGGGAGGTGAAAACGAAAAAAGGTGGGAAAGTGATGACCAAG 1713
             |||:::|||   :::::      :::      |||:::
  804 ......GluAlaLysArgArgArgLysGluGluLysLysThrArg 816
```

```
seq_name: pir2:T22808

seq_documentation_block:
hypothetical protein F56H9.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T22808
R;Burton, J.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19618
A;Accession: T22808
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-770 <WIL>
A;Cross-references: EMBL:Z74473; PIDN:CAA98949.1; GSPDB:GN00023; CESP:F56H9.1
A;Experimental source: clone F56H9
C;Genetics:
A;Gene: CESP:F56H9.1
A;Map position: 5
A;Introns: 235/1; 262/2; 320/1; 367/2; 510/3; 654/1; 681/2


alignment_scores:
           Quality:  162.00              Length:    218
             Ratio:  1.409                Gaps:      9
 Percent Similarity:  52.752   Percent Identity: 27.982

alignment_block:
 US-09-303-518D-463 x T22808   ..

 Align seg 1/1  to: T22808  from: 1  to: 770

      86 ACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCCGAC 135
         |||:::|||:::|||        ||||||   |||   ::::::   ||
      37 ThrValProValSerThrThrThrSerThrSerSerThrThrThrThrTh 53

     136 G......GGAAATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCG 179
         |      ::::::||||||   ||||||:::      ::::::|||||||:
      53 rValProSerThrThrThrSerSerAlaThrThrThrThrValProSerT 70

     180 CAACGGCCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACC 229
         ::|||:::   |||         |||
      70 hrThrThrSerSerThrThrThrThr..................... 78

     230 TGATGATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGC 279
                ::::      |||                 ::::::|||   :::::
      79 .........ThrValProProThrThrThrSerThrThrThrThrThrTh 92

     280 TTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTC 329
         :   |||   |||   |||:::|||         :::||||||:::|||
```

```
 92 rValProProThr...ThrThrSerThrThrThrThrThrThrValProP 108

330 ACATT.........CCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGAT 370
                     |||      |||   |||   |||
108 roThrThrThrSerThrThrThrThrThrThrValProProThrThrThr 124

371 TCAGCCTTTACCGCATCCATT.....................GGGACGGA 399
    |||:::   |||::::::                      :::|||
125 SerSerThrThrThrThrThrThrValProProThrThrThrSerSerThrTh 141

400 TACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTA 449
     |||:::|||:::||||||||||:::   |||:::      ::::::::::
141 rThrThrThrValProProThrThrThrThrSerThrThrThrThrThrV 158

450 TCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCG 499
    ::|||   |||   :::   :::      |||:::||||||
158 alPro...ProThrThrThrSerThrThrThrThrThrThr......... 170

500 TTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAA 549
    :::|||      ::::::||||||      ||||||::::::|||
171 ValProAlaThrThrThrSerThrAlaThrThrThrThrValPro..... 185

550 CGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGT 599
           ||||||:::||||||
186 .......ProThrThrSerThr........................... 190

600 AGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGAT 649
      :::|||   :::::::::||||||   |||:::   :::   |||
191 ..ThrThrThrThrThrValProProThrThrThrSerThrThrThrThr 206

650 CGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAAC 699
      ::::::|||||      :::::::::         ::::::   ||
207 ThrThrValProProThrThrThrThrSerSerThrThrThrThrThrTh 223

700 ATCA 703
    |:::
223 rThr 224
```

seq_name: pir2:T45134

seq_documentation_block:
hypothetical protein [imported] - Microbacterium ammoniaphilum (fragment)
C;Species: Microbacterium ammoniaphilum
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T45134
R;Striebel, H.M.; Seeber, S.; Jarsch, M.; Kessler, C.
Gene 172, 41-46, 1996
A;Title: Cloning and characterization of the MamI restriction-modification system from M
A;Reference number: Z22923; MUID:96257250
A;Accession: T45134
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-529 <STR>
A;Cross-references: EMBL:X79027; NID:g984667; PIDN:CAA55649.1; PID:g1679831
A;Experimental source: ATCC 15354

alignment_scores:

Quality:	161.50	Length:	407
Ratio:	1.009	Gaps:	19
Percent Similarity:	39.312	Percent Identity:	24.079

alignment_block:
US-09-303-518D-463 x T45134 ..

Align seg 1/1 to: T45134 from: 1 to: 529

```
 81 GGCAAACGATCCCTTTATCCGGCAGGTTCT.............CGACC 115
    |||      ::::::   |||   ||||||               |||||
 58 GlyThrGlyThrIleGluProArgGlySerHisTyrArgHisArgArgPr 74

116 GTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGG.........C 156
    |:::           ||||||::::::   |||::::::         |

 74 oAla.........GlnArgArgGlnValHisProLeuGlnArgProHisG 88

157 AGCAGGGGGGA..................................... 167
    ||:::   |||
 88 lnGluProGlyAlaArgGlyGluLeuProValArgAspAspArgAlaGlu 104

168 ............................GCTTGCCNAGCGCAACGGCC 187
                                   |||      |||:::|||
105 ArgGlyArgArgGluProAlaArgProAlaAlaArgAspAlaArgArgHi 121

188 ATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGATT 237
        |||:::|||   |||                           :
121 sLeuArgLeuGlyAlaHis...........................ProA 129

238 CAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGA 287
    ::      |||      |||:::::::|||   |||            |||
129 laArgArgGlyValLeuArgArgHisArgArgHis............Arg 141

288 TCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCCG 337
    :::|||::::::
142 AlaArgGlyGlu..................................... 145

338 ATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATC 387
           |||:::|||      |||   ::::::   |||      ||||
146 .......ArgGlyArgGlyProArgGlnGlnVal...ProArgGlnHisP 159

388 CATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACA 437
    ||   ||||||            :::|||      |||
159 roArgGlyArgArgAspArgAlaGlyArgProGlyLeu............ 171

438 GGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACG 487
      ||||||      |||::::::||||||:::|||      |||:::||
172 HisArgArgArgArgArgAlaArgArgArgGlnGlyArgProGlnValAr 188

488 ACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGC 537
    ||||                                    :::   |
188 gHis.................................GlyAspAspG 193

538 AGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCT 587
    |||||||||:::         ||||||                 :::
193 lnHisArgAlaAspProArgArgProArgAspProArgAlaGlyHisHis 209

588 GACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCG 637
          |||:::      :::   |||      :::|||::::::   ||
210 ProLeuArgGluGlyGlyGlnGlyGlnGluAlaArgProLeuArgAlaAr 226

638 AGCTGGACAGATCGGG................................ 653
    |:::||||||   |||
226 gGlyGlyGlnGlyGlyProGlyGlyProAlaGlyArgHisProAlaLeuG 243

654 .........................CAATGCCGCCGAAGCCTTCAACGG 677
                               |||      |||:::   ::::|||
243 lyAspGluAspArgProArgAlaArgGlnGlyAlaArgAlaAlaHisArg 259

678 CACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAAT...TG 724
    ::::::   :::   :::   ::::::||||||||||||||   :::   ||
260 GluAlaLeuHisLeuArgLeuGlnArgArgArgGlyArgAlaAspGlyCy 276

725 TCGG..............CGCAGGCGATGCCGTGCAGGGTATAAGCGAA 759
    ||||               ||||||      |||||||||
276 sArgAlaGlnGlyArgAlaArgArgAlaArgArgAlaGly.......... 289

760 GGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAA 809
                                  |||:::::::::   |||:::
290 ............................GlyGlyGlyLeuProArgArg 296

810 CAAGAT......GGCGCGCATCAA.......................... 827
    :::|||         ||||||:::
297 GluAspArgValArgAlaHisArgProArgProArgGlyArgGlyGlyAl 313

828 .............CGATTTGGCAGATATGGCGCA...ACTCAAAGACTAT 861
                  |||   |||       ||||||         |||
313 aAlaArgLeuAspArgAlaGlyGlyValGlyAlaArgProAlaArgProA 330
```

```
     862 GCCGCAGCAGCCATCCGCGA.............TTGGGCAGTCCAAAA 896
         |||      |||||||||                 |||:::|||::: 
     330 rgArgLeuArgHisProArgProAlaAspLeuProHisGlyGlyProGln 346

     897 CCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAG 946
           :::|||      |||   :::     :::         |||||
     347 GlyValAlaArgLeuAspHisProGlnGlyLeuGluGlyAlaAlaGlySe 363

     947 CCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGC 996
         |:::   |||                 |||   ::::::|||
     363 rArgArgHis.................ProHisArgLeuArgAlaArgL 374

     997 GGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATT 1046
           |||:::|||   :::                     |||||||||:::
     374 euHisGlnGlyArgGlyAspLeuLeuArgArgProArgArgAspArgLeu 390

    1047 GCCGAAAGGGAAATCCGCCGT 1067
         ::::::|||      |||:::
     391 GlyArgArgGlyProArgGln 397

seq_name: pir2:T45025

seq_documentation_block:
mucin MUC5B, tracheobronchial [imported] - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T45025
R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A;Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
A;Reference number: Z22899; MUID:97166151
A;Accession: T45025
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3570 <DES>
A;Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A;Experimental source: placenta
C;Genetics:
A;Gene: MUC5B


alignment_scores:
            Quality:  158.00              Length:    365
              Ratio:  0.952                Gaps:     16
 Percent Similarity:  45.479   Percent Identity: 25.205

alignment_block:
 US-09-303-518D-463 x T45025   ..

 Align seg 1/1  to: T45025  from: 1  to: 3570

     290 ACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGAT 339
         |||:::|||   |||      |||:::||||||               ||
     570 ThrAlaThrProSerSerThrProGlyThrThrTrp............Il 582

     340 TCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCA 389
         ||||||||||||::::::      |||   :::|||   |||:::|||
     582 eLeuThrLysProThrThrThrAlaThrThrThrAlaSerThrGlySerT 599

     390 TTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGG 439
           :::|||   ::::::|||         ||||||   |||   |||
     599 hrAlaThrAlaSerSerThr...GlnAlaThrAlaGlyThrProHisVal 614

     440 GCGGCGGCTATCCCGCTCCCA..............AAGGCGCGAGGGAT 474
         :::::::::         |||               ||||||      ::
     615 SerThrThrAlaThrThrProThrValThrSerSerLysAlaThrProPh 631

     475 ATATACAGCTACGACATAAAAGGCGTTGCCCAAAATA...TCCGCCTCAA 521
         :   :::      |||      |||||||||   :::   |||::::::|
     631 eSerSerProGlyThrAlaThrAlaLeuProAlaLeuArgSerThrAlaT 648

     522 CCTGACCGACA............ACCGCAGCACCGGACAACGGCTTGCCG 559
         ||   ||||||           ||||||   |||::::::::::|||
     648 hrThrProThrAlaThrSerPheThrAlaIleProSerSerSerLeuGly 664

     560 ACCGTT..................TCCACAATGCCGGCGCTATGCTGACG 591
         ||||::                    :::|||   |||
     665 ThrThrTrpThrArgLeuSerGlnThrThrThrProMet.......... 677

     592 CAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCT 641
                 ||||||   |||:::|||   |||::::::::::|||:::
     678 .........AlaThrMetSerThrAlaThrProSerSerThrProGluT 691

     642 GGACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCG 691
           :::                          ||||||:::|||      :::
     691 hrVal..................HisThrSerThrValLeuThrThrThr 701

     692 TCAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCC 741
         :::   |||::::::|||      :::      :::   ::::::   ||
     702 AlaThrThrThrGlyAlaThrGlySerValAlaThrProSerSerThrPr 718

     742 GTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGG 791
         |      :::   :::                  :::   |||:::::::
     718 oGlyThrAlaHisThrThrLysValLeuThrThrThrThrThrGlyPheT 735

     792 TCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATA 841
         :::::   |||   ::::::      ||||||   |||:::      :::
     735 hrAlaThrProSerSerSerProGlyArgAlaArgThrLeu...ProVal 750

     842 TGGCGCAACTCAAAGACTATGCCGCAG.................CAGCC 873
         |||   ::::::   |||   |||                       ||
     751 TrpIleSerThrThrThrThrProThrThrArgGlySerThrValThrPr 767

     874 ATCCGCGATT...GGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGA 920
         ||||:::|||   |||   :::   ||||||          :::
     767 oSerSerIleProGlyThrThrHisThrProThrValLeuThrThrThrT 784

     921 AGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAG 970
               ::::::   :::      :::   :::||||||::::::   :::
     784 hrThrThrValAlaThrGlySerMetAlaThrProSerSerSerThrGln 800

     971 CTGTCCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAG 1020
           |||
     801 ThrSer..................................... 802

    1021 CGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAG 1070
                                        |||:::|||||||||
     803 ..............................GlyThrProProSerL 808

    1071 CGACAATTTTGCCGATGCGGCATACG.....................CCA 1099
           |||                    |||                   |||
     808 euThrThrThrAlaThrThrIleThrAlaThrGlySerThrThrAsnPro 824

    1100 AATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAG 1149
         ::::::   |||   |||   |||          |||   |||   :::::
     825 SerSerThrProGlyThrThrProIleProProValLeuThrThrThrAl 841

    1150 CGTTACGGCAAAGAAAACATCA........................... 1171
         ::::|||         ||||||
     841 aThrThrProAlaAlaThrSerSerThrValThrProSerSerAlaLeuG 858

    1172 .............CCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATG 1207
                       ||||||   |||          :::|||:::
     858 lyThrThrHisThrProProValPro.........AsnThrThrAlaThr 871

    1208 TCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTT 1250
         :::::: ||||||          |||::::::   :::||||||
     872 ThrHis.GlyArgSerLeuSerProSerSerProHisThrVal 885

seq_name: pir2:S57180

seq_documentation_block:
probable membrane protein YJR151c - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein J2223; serine/threonine-rich protein YJR151c
C;Species: Saccharomyces cerevisiae
```

```
C;Date: 23-Aug-1995 #sequence_revision 08-Sep-1995 #text_change 05-Nov-1999
C;Accession: S57180
R;Scarcez, T.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S57169
A;Accession: S57180
A;Molecule type: DNA
A;Residues: 1-1161 <SCA>
A;Cross-references: EMBL:Z49651; NID:g1015902; PID:g1015903; GSPDB:GN00010; MIPS:YJR151c
C;Genetics:
A;Gene: MIPS:YJR151c
A;Map position: 10R
C;Keywords: transmembrane protein


alignment_scores:
           Quality:  153.50            Length:    429
             Ratio:  0.717              Gaps:     12
 Percent Similarity:  49.883   Percent Identity: 22.611

alignment_block:
 US-09-303-518D-463 x S57180   ..

 Align seg 1/1  to: S57180  from: 1  to: 1161

      77 ATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTC 126
         |||:::   :::||||||   |||         :::|||::::::   ||
     117 IleTyrThrAlaIleProThrSerThrSerThrThrThrThrLysSerSe 133

     127 GAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNA 176
         |:::   |||      ||||||      ::::::      |||      |
     133 rThrSerThrThrProThrThrThrIleThrSerThrThrSerThrThrS 150

     177 GCGCAACGGCCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCC 226
         ||:::|||      ::::::      :::   :::
     150 erThrThrProThrThrSerThrThrSerThrThrProThrThrSerThr 166

     227 ACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTC 276
         |||                                 ::::::|||   ||
     167 Thr.................SerThrThrProThrThrSerThrThrSe 177

     277 CGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGC 326
         |:::   |||   |||:::|||:::|||      |||:::|||:::
     177 rThrThrProThrThrSerThrThrSerThrThrProThrThrSer.... 192

     327 CTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCC 376
                       |||      ::::::|||   |||::::::::::
     193 .................ThrThrSerThrThrProThrThrSerThrThr 203

     377 TTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTAT 426
           |||:::         :::   |||:::|||   |||   |||:::
     204 SerThrThrProThrThrSerThrThrSerThrThrProThrThrSerTh 220

     427 GACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATAT 476
          |||:::         ::::::         |||
     220 rThrSerThrThrProThrThrSerThrThrProThr............. 232

     477 ATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGA 526
           |||:::||||||      :::            :::   ||||||
     233 ..ThrSerThrThrSerThrThrSerGlnThrSerThrLysSerThrThr 248

     527 CCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCC 576
         ||||||||:::::::   ::::::      ||||||:::||||||   ||
     249 ProThrThrSerSerThrSerThr...ThrProThrThrSerThrThrPr 264

     577 GGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCG 626
         |:::                      :::|||   |||:::||||||
     264 oThrThr.................SerThrThrSerThrAlaProThrT 275

     627 ATACAGCCCCGAGCTGGACAGATCGG............GCAATGCCGCCG 664
           ::::::   |||   |||:::                :::      |||
     275 hrSerThrThrSerThrThrSerThrThrSerThrIleSerThrAlaPro 291

     665 AAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCA 714
               ||||||:::   :::      |||   :::|||::::::   ::
     292 ThrThrSerThrThrSerSerThrPheSerThrThrSerSerAlaSerAlaSe 308

     715 GGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTC 764
         ::::   :::|||:::   |||         :::   |||:::
     308 rSerValIleSerThrThrAlaThrThrSerThrThrPheAlaSerLeuT 325

     765 AAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGA 814
           |||      :::   ||||||                     |||
     325 hrThrProAlaThrSerThrAla.........SerThrAspHisThrThr 338

     815 TGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCC 864
               :::||||||         :::      ::::::   |||
     339 SerSerValSerThrThrAsnAlaPheThrThrSerAlaThrThrThrTh 355

     865 GCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGG 914
         :::   :::   |||::::::|||:::      :::      ::::
     355 rThrSerAspThrTyrIleSerSerSerSerProSerGlnValThrSerS 372

     915 CATAGAAGCCGTCAGCAATA............TCTTTATGGCAGCCATCC 952
         ::   :::|||:::::::::            |||      :::|||:::
     372 erAlaGluProThrThrValSerGluValThrSerSerValGluProThr 388

     953 CCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCG..... 997
           |||:::   :::      |||:::|||   |||:::   :::
     389 ArgSerSerGlnValThrSerSerAlaGluProThrThrValSerGluPh 405

     998 ......................GCATCACGGCACATCCTGTCAAGCGGTC 1025
                                 |||   ::::::   ||||||::::
     405 eThrSerSerValGluProThrArgSerSerGlnValThrSerSerAlaG 422

    1026 GCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACA 1075
         ::      :::   |||:::      ::::::      |||:::   :::
     422 luProThrThrValSerGluPheThrSerSerValGluProThrArgSer 438

    1076 ATTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGA 1125
                    :::   |||   ::::::::::|||   |||:::   ||
     439 .........SerGlnValThrSerSerAlaGluProThrThrValSerGl 452

    1126 AATA......TCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACAT 1169
         |:::      ||||||:::         |||      :::   :::::::
     452 uPheThrSerSerValGluProThrArgSerSerGlnValThrSerSerA 469

    1170 CACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAG 1219
         ::   |||            :::   |||::::::::::::      :::
     469 laGluProThrThrValSerGluPheThrSerSerValGluProThrArg 485

    1220 ACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCG 1269
         ::::::   :::      :::|||         ||||||::   |||
     486 SerSerGlnValThrSerSerAlaGluProThrThrValSerGluPhe.. 501

    1270 AATTTTGAGAAGCACGTGAAATATGATACGAAGCTCG 1306
                   |||:::   :::   ||||||||||||
     502 .......ThrSerSerValGluProIleArgSerSer 511

seq_name: pir2:JQ0405

seq_documentation_block:
hypothetical 119.5K protein (uvrA region) - Micrococcus luteus
N;Alternate names: ORF 1 protein
C;Species: Micrococcus luteus, Micrococcus lysodeikticus
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 04-Feb-2000
C;Accession: JQ0405
R;Shiota, S.; Nakayama, H.
Mol. Gen. Genet. 217, 332-340, 1989
A;Title: Micrococcus luteus homolog of the Escherichia coli uvrA gene: identification
A;Reference number: S04781; MUID:89364717
A;Accession: JQ0405
A;Molecule type: DNA
A;Residues: 1-1106 <SHI>
A;Cross-references: EMBL:X15867
A;Note: this reading frame extends between two stop codons and does not begin with a
```

```
A;Note: the gene encoding this protein overlaps uvrA gene
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;


alignment_scores:
            Quality:  150.00              Length:    550
              Ratio:  0.708                Gaps:     33
 Percent Similarity:  38.545   Percent Identity: 24.364

alignment_block:
 US-09-303-518D-463 x JQ0405   ..

 Align seg 1/1  to: JQ0405  from: 1  to: 1106

      51 CCTGCCGATGCATGCACACGCCTCAGATTTGGCAAACGATCCCTT..... 95
         |||               |||||||||   |||:::|||   |||
     601 ProProGlyHisThrAspArgLeuArgAlaGlyArgArgProLeuArgPr 617

      95 ..................................................... 95

     617 oArgArgAlaValHisArgProAlaProAlaGlyGlnProProProHisA 634

      96 ..........TATCCGGCAGGTTCTCGA......CCGTCAGCATTTCGA 128
                   |||||||||   |||      |||         |||
     634 rgAspProProAlaProAlaGlyProArgGlnHisProHisArgArgArg 650

     129 ACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGC 178
         :::||||||                  ::::::||||||         ::
     651 AlaArgArgGlyHisAsp.........ArgArgGlyGlyLeuAspArgGl 664

     179 GCAACGGCCATATCGGATTGGGAAACATACAAAG......CCATCAGTTG 222
         ::::||||||   |||   |||:::      :::         ::::::|
     664 yHisArgProSerArg...GlyArgValArgArgArgGlyArgAlaLeuG 680

     223 GGCCACCTGATGATTCAACAGGCGGC.........CGTTGAAGGAAATAT 263
         ||   |||      |||   |||            |||   |||:::
     680 lyLeuProGlyGlySerGlnGlyGluHisAlaValArgHisArgArgLeu 696

     264 CGGCTACATTGTCCGCTTTTCCGA.......................... 287
            |||:::   ||||||   |||
     697 ProLeuArgProProLeuHisArgGlyAlaGlyAlaAlaSerArgProGl 713

     288 .......TCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCA 330
                 :::||||||:::   |||         |||:::||| ::::::
     713 yGluGlyAlaArgAlaAspGlyPro.........ArgArgPro.GlyGlu 726

     331 CATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGA........... 369
         :::   ::::::      :::   :::|||:::   |||
     727 GlnProGluGlyArgLeuGlyProGlyProAlaArgGlyProHisGlyAr 743

     370 ..........TTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACC 409
                    |||   ::::::|||
     743 gAspGlyArgValArgLeuArgGlnValHis.................... 753

     410 ATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCC 459
             ||||||      |||      ||||||   |||   ||||||  |
     754 .....AlaAspGlnArgAspProLeuGlnGlyProGlyGlnProAla..G 768

     460 AAAGGCGCGAGGGATATATACAG.........CTACGACATAAAAGGCGT 500
         ||||||||::::::      :::         :::|||         |||
     768 lnArgArgGlnAlaArgAlaArgProAlaProValArgGlyGlyThrArg 784

     501 TGCCCA............AAATATCCGCCTCAACCTGACCGA........ 530
         :::|||                     |||:::|||   |||
     785 AlaProGlyGlnGlyGlyProArgGlyProGluProHisArgAlaHisAl 801

     531 .......CAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAAT 573
                 ||||||:::|||              |||      ||||||:
     801 aThrLeuGlnProArgHisLeuHisGlyArgValArgArgAspProGlnA 818

     574 GC..................................................... 575
         ::
     818 laLeuArgGlyAspProArgGlyGluGlyProGlyLeuProAlaGlyPro 834

     576 .....................................CGGCG 580
                                              |||||
     835 ValLeuLeuGlnHisGlnGlyArgAlaLeuArgGlyValArgGlyArgAr 851

     581 CTATGCTGACGCAAGGAGTAGG.................CGACGGATTC 612
         |:::|||:::   |||::::::                |||   ::::
     851 gHisAlaGluAspArgAspGluLeuProAlaGlyArgLeuArgAlaValA 868

     613 AA......ACGCGCCACCCGATACAGCCC.........CGAGCTGGACAG 647
         ::         |||   |||:::||||||         |||:::
     868 rgGlyValProArgGlyProValGlnProGlyAspAlaArgGlyHisLeu 884

     648 ATCGGGCAATGCCGCCGAAGC..................... 668
            ||||||   :::|||
     885 GlnGlyGlnGluHisArgArgGlyProArgHisAlaAspArgGlyGlyAr 901

     669 .......CTTCAACGG.........................CACTGCA 684
               |||||||||                         |||:::|
     901 gGlyLeuLeuGlnArgValHisProHisLeuAlaValProGlyHisAlaA 918

     685 GATATCGTCAA.........................AAACATCATCGG 707
         ||   |||:::                        :::||||||
     918 rgArgArgArgSerGlyLeuArgProSerGlyProAlaArgHisHisAla 934

     708 CGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAG.. 755
            |||||||||:::      |||         :::||||||   :::
     935 LeuGlyArgArgGlyProAlaArg.........GluAlaGlyGlyArgAl 948

     756 ....CGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCC 801
             :::   |||::::::      |||   |||   |||   |||   |
     948 aProGluAlaLeuGlnArgProHisHisLeuArgAlaGlyArgAlaAspH 965

     802 AC..................................................... 803
         ||
     965 isGlyValAlaLeuArgArgHisProGlnAlaProAlaArgAlaSerVal 981

     804 .........CGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGG 844
                 :::   |||   |||||||||:::|||
     982 ProArgGlyGlnGlyGlnHisGlyAlaHisHisArg.............. 993

     845 CGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAA 894
          |||      |||      ::::::      ||||||         |||:
     994 .AlaGlnProArgArgAspGlnGluArgGlyProArg...AspArgProA 1009

     895 AACCC......CAATGCCGCACAAGGCAT..................... 917
         ::|||      :::   |||   ||||||
    1009 rgProGlyGlyArgLeuArgArgArgHisAspArgGlyHisGlyHisAla 1025

     918 .........AGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCA 958
                 |||:::|||::::::      :::|||  |||   |||
    1026 GlyGlyGlyArgAlaArgArgArgGluProHisGly..ProValProArg 1041

     959 AAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCATCACGGCA 1008
                       ::::::   ::::::||||||:::         ::
    1042 ..............GlyAlaAlaArgValAlaTrpThrProProThrGl 1053

    1009 CATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAA 1058
         :      ::::::|||:::   |||   |||   :::          |||
    1053 nAspGluAspGluGlyGluProTrpArgArgArgArgProCysAlaGlyA 1070

    1059 ATCCGCCGTCAGCGACAATTTTGCCGATG......CGGCATACGCCAAAT 1102
           ||||||   :::         |||         |||   :::   :::
    1070 laProProArgCysCysGlySerProAlaValProArgSerSerAspGly 1086

    1103 ACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGT 1152
         :::||||||         |||
    1087 AlaArgProArgGluAlaPro............................ 1093

    1153 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACG 1198
                 ::::::|||||||||      ||||||   :::|||
    1094 .........SerSerSerProProThrThrCysArgSerSerThr 1105
```

```
seq_name: pir2:T02345

seq_documentation_block:
hypothetical protein KIAA0324 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999
C;Accession: T02345
R;Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;
re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
submitted to the EMBL Data Library, March 1998
A;Description: Sequencing of human chromosome 16p13.3.
A;Reference number: Z14664
A;Accession: T02345
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1791 <RIC>
A;Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AAC08453.1; PID:g2996650
C;Genetics:
A;Map position: 16
A;Introns: 1610/2; 1706/2
A;Note: KIAA0324


alignment_scores:
           Quality:  150.00            Length:    598
             Ratio:  0.558              Gaps:      28
 Percent Similarity:  44.983   Percent Identity: 22.742

alignment_block:
 US-09-303-518D-463 x T02345   ..

 Align seg 1/1  to: T02345  from: 1  to: 1791

      39 ACTGGCAGTGTGCCT...GCCGATGCATGCACACGCCTCAGATTTGGCAA 85
         :::|||||||   |||   :::|||:::   :::|||||||         ::
     611 SerGlySerSerProGluValAspSerLysSerArgLeuSer...ProAr 626

      86 ACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCCGAC 135
         :||||||         :::|||||||   |||::::::           |
     626 gArgSer......ArgSerGlySerSerProGluValLysAspLysProA 641

     136 GGGAAATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGG 185
         ||      |||      |||   |||:::|||:::   ::::::   :::
     641 rgAlaAlaPro......ArgAlaGlnSerGlySerAspSerSerProGlu 655

     186 CCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGA 235
         |||         :::   ::::::   |||      :::         ::
     656 ProLysAlaProAlaProArgAlaLeuProArgArgSerArgSerGlySe 672

     236 TTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCC 285
         :|||:::   ||||||                              |||  :
     672 rSerSerLysGlyArg.....................GlyProSerProG 682

     286 GATCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTC 335
         :::::   ::::::         :::   :::::::|||
     682 luGlySerSerSerThrGluSerSerProGluHisProPro......... 695

     336 CGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCA 385
                 :::||||||   :::|||         :::      |||::
     696 .........LysSerArgThrAlaArgArgGlySerArgSerSerProGl 709

     386 TCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCA 435
         :|||      :::   |||||||||   |||||||||      |||::::
     709 uProLysThrLysSerArgThrProProArgArgArgSerSerArgSerS 726

     436 CAGGGCGG..............CGGCTATCCCGCTCCCAAAGG...... 464
         ::   :::               ||||||||||||   :::|||
     726 erProGluLeuThrArgLysAlaArgLeuSerArgArgSerArgSerAla 742

     465 ..............CGCGAGGGATATATACAGCTACGACATAAAAGGCG 499
                        ::::::           ||||||:::|||
     743 SerSerSerProGluThrArgSerArgThrProProArgHisArgArgSe 759

     500 TTGCCCAAAATATCCGCCTCAACCTGACCGACAA..............  533
                        |||:::|||   ::::::
     759 rProSerValSerSerProGluProAlaGluLysSerArgSerSerArgA 776

     534 ...................CCGCAGCACCGGACAACGGCTTGCCGA...  560
                           |||:::   :::|||||:::   |||
     776 rgArgArgSerAlaSerSerProArgThrLysThrThrSerArgArgGly 792

     561 ......CCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGG..  602
               |||   |||:::   |||           :::||||||||||
     793 ArgSerProSerProLysProArgGlyLeuGlnArgSerArgSerArgSe 809

     603 .CGACGGATTCAAACGCGCCACCCG...ATACAGCCCCGAGCTGGACAGA 648
          ||||||   :::|||||||           :::   |||:::|||
     809 rArgArgGluLysThrArgThrThrArgArgArgAspArgSerGly.... 824

     649 TCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAA 698
                        |||   |||          |||   ||||||:::
     825 .................SerSerGlnSerThrSerArgArgArgGlnArg 835

     699 CATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGG 748
            :::   |||:::|||   :::   ||||||||||||      :::||
     836 SerArgSerArgSerArgValThr...ArgArgArgArgGlyGlySerGl 851

     749 GTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTT 798
         ||||                        |||:::|||
     851 yTyr.......................HisSerArg.............. 855

     799 TCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCA 848
                      :::   |||:::|||:::   :::|||   ::::::
     856 ..............SerProAlaArgGlnGluSerSerArgThrSerSer 867

     849 ACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACC 898
            :::|||      ||||||      |||                   ||
     868 ArgArgArgArgGlyArgSerArgThrPro.................Pr 878

     899 CCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCC 948
         |      |||   |||   |||||||||                    :::
     878 oThrSerArgLysArgSerArgSerArg.................ThrS 889

     949 ATCCC............CATCAAAGGGATTGGAGCTGTCCGGGGAAAATA 986
           |||           ::::::|||:::   :::   |||:::
     889 erProAlaProTrpLysArgSerArgSerArgAlaSerProAlaThrHis 905

     987 CGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCG 1036
         |||   :::|||         :::            ::::::      ||
     906 ArgArgSerArg.........SerArgThrProLeuIleSerArgArgAr 919

    1037 CGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGAT 1086
         |:::|||               :::   |||:::|||:::      |||
     919 gSerArgSerArgThrSerProValSerArgArgArgSerArgSerArgT 936

    1087 GCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTC 1136
           ::::::
     936 hrSerVal.......................................... 938

    1137 AAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGC 1186
                        |||:::|||:::::          :::   ::
     939 ..............ThrArgArgArgSerArgSerArgAlaSerProVa 950

    1187 CGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACCCGAAG 1236
         ::::   :::|||::::::   :::             |||||||||
     950 lSerArgArgArgSerArgSerArg............ThrProProValT 963

    1237 ACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAATTTTGAGAAGCACGT 1286
           ||||||         |||   |||::: |||:::   :::::::::
     963 hrArgArgArgSer...ArgSerArgThr.ProThrThrArgArgArgSe 978

    1287 GAAATATGATACGAAGCTCGATATTCAAGAATTATCGGGGGGCGGTATAC 1336
          :::      |||            ::::::   |||   :::      |
     978 rArgSerArgThrProProValThrArgArgArgSerArgSerArgThrP 995

    1337 CTAAGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAGGTTGATAGG 1386
```

```
      ||         ||||||         :::   |||
   995 ro.........ProValThrArgArgArgSerArg............Ser 1004

  1387 AAGCTTAATAAATTGACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGA 1436
       :::   :::   :::|||   |||::::::   :::::::::
  1005 ArgThrSerProIleThrArgArgArgSerArgSerArgThrSerProVa 1021

  1437 AACGAGAAGAAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAG 1486
       ||||||||||||||:::|||            :::      :::|||
  1021 lThrArgArgArgSerArgSerArgThrSerProValThrArgArgArg. 1037

  1487 AATGGGAAAATAAAACAGGGTTAGATTTTAATCATTTTATAGGTGGTGAT 1536
                                                     :::
  1038 .............................................Ser 1038

  1537 ATCAATAAGAAAGGCACAGTAACAGGAGGGCATAGTCTAACCCGTGGTGA 1586
          ::::::   :::   ||||||      :::|||   :::|||
  1039 ArgSerArgThrSerProValThrArgArgArgSerArgSerArgThrPr 1055

  1587 TGTACGGGTGATACAACAAACCTCGGCA..............CCTGATA 1621
               :::   :::::::::   :::               |||   |
  1055 oProAlaIleArgArgArgSerArgSerArgThrProLeuLeuProArgL 1072

  1622 AACATGGGGTTTATCAAGCGACAGTGGAAATTAAAAAGCCTGATGGAAGT 1671
       ||::::::      ::::::   :::   |||::::::   :::   |||
  1072 ysArgSerArgSerArgSerProLeuAlaIleArgArgArgSerArgSer 1088

  1672 TGGGAGGTGAAAACGAAAAAAGGTGGGAAAGTGATGACCAAG 1713
               :::|||   :::|||   :::   :::||||:::
  1089 ArgThrProArgThrAlaArgGlyLysArgSerLeuThrArg 1102

seq_name: pir2:JW0067

seq_documentation_block:
chitinase (EC 3.2.1.14) A - Emericella nidulans
N;Alternate names: chiA
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 11-Jan-2002
C;Accession: JW0067
R;Takaya, N.; Yamazaki, D.; Horiuchi, H.; Ohta, A.; Takagi, M.
Biosci. Biotechnol. Biochem. 62, 60-65, 1998
A;Title: Cloning and characterization of a chitinase-encoding gene (chiA) from Aspergill
A;Reference number: JW0067; MUID:98162139
A;Accession: JW0067
A;Molecule type: mRNA
A;Residues: 1-660 <TAK>
A;Cross-references: DDBJ:D87895; NID:g2821948; PID:g2828335
C;Comment: This enzyme hydrolyzes chitin at belta-1,4 bonds between N-acetyglucosamine r
C;Genetics:
A;Gene: chiA
C;Keywords: glycosidase; hydrolase


alignment_scores:
            Quality:  149.00            Length:    399
              Ratio:  0.819              Gaps:     19
 Percent Similarity:  45.614   Percent Identity: 23.559

alignment_block:
 US-09-303-518D-463 x JW0067   ..

 Align seg 1/1  to: JW0067  from: 1  to: 660

      97 ATCCGGCAGGTT.........CTCGACCGTCAGCATTTCGAACCCGACGG 137
         ||||||::::::         ||||||   :::|||       ||||||
     332 IleArgGluIleLeuTyrAspLeuAspProAsnHis.ProProProThrT 348

     138 GAAATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCC 187
           :::   |||   :::   :::             ||||||:::|||:::
     348 hrSerProThrProThrProThr............ProSerThrThrThr 360

     188 ATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGATT 237
           |||      :::|||

     361 ThrSerThrThrSerThr............................... 366

     238 CAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGA 287
                                      :::::::|||   ||||||
     367 ........................ThrSerThrThrSerAlaThrSerT 375

     288 TCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCCG 337
         |||:::|||:::|||         :::|||:::
     375 hrThrSerThrThrSerThrThrSerThrThrSer.............. 386

     338 ATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATC 387
                         :::::::|||   |||::::::::::   |||:::||
     387 ..............ThrThrProThrThrSerThrThrSerThrThrSe 398

     388 CATTGGGACGGATACG...AACACCATCCCGCCGACGGCTATGACGGGCC 434
         |     |||   |||   :::       |||   ||||||   ::::::
     398 rThrThrThrProThrProSerProSerProSerThrAlaSerSerSerT 415

     435 ACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCT 484
               ::::::   |||   ||||||
     415 hrThrGluThrValThrProSerProLysProSer.............. 426

     485 ACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAAC 534
                         |||:::   |||:::::::|||      |||::
     427 ..................ProSerGluSerSerThrThrSerGluThrSe 437

     535 CGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTAT 584
         ::::   |||::::::::::   |||:::||||||      |||:::
     437 rSerLeuProSerThrSerThrProValValSerGluThrProSer.... 452

     585 GCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCC 634
               :::   :::      |||:::|||:::|||||||||   ::::::
     453 .....GluThrLysThrProThrSerSerSerAlaProProLeuSerSer 467

     635 CCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCA 684
           |||                     |||      :::::::::
     468 SerSer.....................ProValGlyGlySerSerSerTh 477

     685 GATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGG 734
           ||||||:::||||||:::      :::         ::::::::::
     477 rAlaSerSerSerThrSerThrProSerGluThrProSerThrSerSerT 494

     735 CGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACG 784
         ::                |||   ::::::|||              :::
     494 hrArg..............AlaValSerGluThrSerThrHisIleSer 505

     785 GCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTG 834
         :::   :::         |||:::|||         :::|||:::
     506 ThrSerThrSerSerGlyProGluThrSerLeuThrGlySerSerThr.. 521

     835 GCAGATATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTG 884
          ::::::      :::|||:::::::::|||:::      |||||||||:
     522 .SerValProAlaThrSerSerSerValProSer......SerAlaIleS 536

     885 GGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCG......... 925
         ::   |||:::||||||:::      ::::::   :::|||
     536 erProSerSerThrProValIleSerGluThrProArgProProValThr 552

     926 ...TCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT 972
           |||:::   |||   :::      |||   |||   :::   :::
     553 SerSerSerSerSerThrPheValSerSerThrSerThrSerThrAspCy 569

     973 GTCCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCG 1022
          |||   ::::::|||      |||   :::   |||      ||||||:
     569 sSerGluSerSerThr......AlaIleGlyThrHis......SerSerS 582

    1023 GTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCG 1072
         ::      |||                         :::   |||||||||
     582 erSerIleTrp.......................GluThrProSerAla 590

    1073 ACAATTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCC 1122
         :::                  ||||||   :::   |||   |||   ||
     591 Ser.................ThrProAlaAlaSerProSerThrSerPr 601
```

```
   1123 CGAAATATCCGTTCAAACTTGGAGCAGCGTTACG..............  1156
        ||||            |||   :::::::::|||
    601 oGlu...........ThrThrLysThrLeuThrValPheProArgArgA  614

   1156 ..................................................  1156

    614 laProProCysLeuProGluProProProArgProLeuCysProAlaLeu  630

   1157 ............GCAAAGAAAACATCACCTCCTCAACCGTGCCGC  1189
                    |||      |||      ||||||||||||:::
    631 PheProGlnLeuAlaValAlaThrArgLysProGlnProCysGln  645

seq_name: pir2:T30944

seq_documentation_block:
surface protein precursor - Enterococcus faecalis
C;Species: Enterococcus faecalis
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T30944
R;Shankar, V.; Baghdayan, A.S.; Huycke, M.M.; Lindahl, G.; Gilmore, M.S.
Infect. Immun. 67, 193-200, 1999
A;Title: Infection-derived Enterococcus faecalis strains are enriched in esp, a gene enc
A;Reference number: Z20943; MUID:99081742
A;Accession: T30944
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1873 <SHA>
A;Cross-references: EMBL:AF034779; NID:g3873186; PID:g3873187; PIDN:AAD09858.1


alignment_scores:
           Quality:  148.00            Length:    695
             Ratio:  0.468              Gaps:     36
 Percent Similarity:  45.468   Percent Identity: 21.871

alignment_block:
 US-09-303-518D-463 x T30944   ..

 Align seg 1/1  to: T30944  from: 1  to: 1873

     80 TGGCAAACGATCCCTTTATCCGGCAGGTTCTCG.............  112
        |||:::|||   |||   :::::::::   :::
    604 TrpLysThrGluProAspThrSerAsnValThrAspSerThrLysGlyIl  620

    113 ....ACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAG  158
            ||||||:::|||:::|||   |||   ::::::      :::   |
    620 eValThrValLysIleGlyAsnArgThrPheAspValAspValGluPheA  637

    159 CAGGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATAC  208
        ||         :::   ::::::   |||:::::::::      ||||||
    637 la.........ValLysAlaSerGlnAlaMetGluAsnAspAlaThrTyr  650

    209 AAAGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGA  258

    650 ..................................................  650

    259 AATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTC  308
                                |||||||||   |||
    651 .....................ValProIleThrThrThr.........  656

    309 GCCCTTCGACAACCATGCCTCAC...ATTCCGATTCTGACGAAGCCGGTA  355
          |||:::||||||:::             |||   :::   ||||||
    657 ..ProGluThrThrIleGlnSerGlyLysProThrPheAspLysProAsp  672

    356 GTCCCGTTGACGGATTCAGCCTTTACC.....................  382
        |||||||||:::::::::||||||:::
    673 ValProLeuAlaAsnAspAlaPheSerValLeuAspValTyrAsnLysAs  689

    383 .........GCATCCATTGGGACGGATACGAAC......ACCATCCCGC  416
                |||||||:::   ::::::|||:::      |||:::   |
    689 pPheGlyAsnAlaSerValAspAlaAsnThrGlyIleValThrPheThrP  706

    417 CGACGGCTATGACGGGCCACAGG.........GCGGCGGCTATCCCGCTC  457
        ||:::      :::|||:::            :::::::::|||||:::
    706 roAlaLysGlyValGlyGluSerGluProIleThrGlyThrIleProIle  722

    458 CCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAA  507
          |||   :::   :::      |||                ::::|||
    723 ..LysIleValTyrGlnAspGlySerValGlyThrThrAspLeuAlaVal  738

    508 AATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGC  557
        ::::::      |||:::   :::|||         |||:::::::::
    739 ThrValSerLysAsnIleTyrGluAsn......ProGlyGluAsnIlePr  753

    558 CGACCGTTTCCACAATGCC.................GGCGCTATGCTGA  589
              :::|||::::::                  |||:::   :::
    753 oAlaGlyTyrHisLysValThrPheThrAlaGlyGluGlyThrSerIleG  770

    590 CGCAAGGA..............GTAGGCGACGGATTCAAACGCGCCACC  624
          :::|||               |||   ||||||    :::   |||
    770 luSerGlyThrThrValPheAlaValLysAspGlyVal.SerLeuProGl  786

    625 CGATACAGCCCCG.............................AGCT  641
         |||      |||                             :::|
    786 uAspLysLeuProValLeuLysAlaLysAspGlyTyrThrAspAlaLysT  803

    642 GGACAGATCGGGCAATGCCGCCGAA..............GCCTTCAAC  675
        ||   :::::::|||      |||                  |||
    803 rpProGluGluAlaThrGlnProIleLysAlaAspAspThrGluPheVal  819

    676 GGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAATTGT  725
        ::::::|||      :::   :::||||||         |||:::   ::
    820 SerSerAlaThrLysLeuAspAspIleIleGluAsnProGlyAspAsnIl  836

    726 CGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTG  775
        :   ||||||            ::::::   |||::::::   ::::
    836 eProAlaGlyTyrHisLysValThrPheThrAlaGlyGluGlyThrSerI  853

    776 TCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATC  825
        ::      |||      ::::::::::::::::::   ::::::
    853 leGluSerGlyThrThrValPheAlaValLysAspGlyValSerLeuPro  869

    826 AACGATTTGGCAGATATGGCGCAACTCAAAGAC...TATGCCGCAGCAGC  872
        :::|||         :::   :::   ||||||   |||:::   |||
    870 GluAspLysLeuProValLeuLysAlaLysAspGlyTyrThrAspAlaLy  886

    873 CATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAG  922
              :::   |||:::|||   |||      |||:::      |||
    886 sTrpProGluGluAlaThrGln...ProIleLysAlaAspAspThrGluP  902

    923 CCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGA...  969
        ||||||:::                        |||::::::   |||
    902 heValSerSerAlaThrLysLeuAspAspIleIleGluAsnProGlyAsp  918

    970 ..................................GCTGTCCGGGGAAA  983
                                           |||   :::|||
    919 AsnIleProAlaGlyTyrHisLysValThrPheThrAlaGlyGluGlyTh  935

    984 ATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGG  1033
              :::|||   |||:::      ||||||
    935 rSerIleGluSerGlyThrThrValPheAlaValLys.............A  948

   1034 GCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTC.........AGCGAC  1074
        :::::::::||||||:::   |||      |||         :::|||
    948 spGlyValSerLeuProGluAspLysLeuProValLeuLysAlaLysAsp  964

   1075 AATTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCG  1124
        ::::::::::|||         ||||||:::|||:::
    965 GlyTyrThrAsp.........AlaLysTrpProGly.............  973

   1125 AAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCT  1174
                                                        :
    974 ..........................................G  974

   1175 CCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAA  1224
```

```
     :::::|||   |||                  :::|||      |||      .
 974 luAlaThrGlnPro.................IleLysAlaAspAspThr 984

1225 CGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAATTT 1274
     :::     ::::::::::::   :::|||   |||:::      :::
 985 GluPheValSerSerAlaThrLysLeuAspAspLysSer......AspAl 999

1275 TGAGAAGCACGTG............AAATATGATACGAAGCTCGATATTC 1312
      :::|||::::::           |||      |||:::|||:::   :
 999 aAspLysTyrThrProGluGlyGlnLysValThrThrGluLeuAsnLysG 1016

1313 AAGAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTG...TTTGAT 1359
     ::      ::::::   ||||||   :::   |||   :::      |||
1016 luProAspAlaSerGluGlyIleLysAsnLysLysAspLeuProLysAsp 1032

1360 GCGAAACCGAGATGG............GAGGTTGATAGGAAGCTTAATAA 1397
     ||||||     |||              :::::::::          |||||
1033 AlaLysTyrThrTrpLysGluLysValAspIleSerThrAlaGlyAsnLy 1049

1398 ATTGACAACT.........................CGTGAGCAGGTGG 1420
     |      |||                        :::::::||||
1049 sLysGlyThrValValValThrTyrSerAspGlySerSerAspGluValG 1066

1421 AGAAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGAGTAGTCAGTTTAAA 1470
     ||   :::|||      |||   :::||||||:::::::::::::::::
1066 luValAspValThrValThrAspAsnArgSerAspAlaAspLysTyrGlu 1082

1471 GCCCATGCGCAACGAGAATGGGAAAATAAAACAGGGTTAGATTTTAATCA 1520
          ::::::   |||     :::
1083 ProThrValGluGlyGluLysValGlu..................... 1091

1521 TTTTATAGGTGGTGATATCAATAAGAAAGGCACAGTAACAGGAGGGCATA 1570
         |||||||||   ::::::         :::||||||:::      :
1092 ....IleGlyGlyLysValAspLeuThrAspAsnValThrAsnLeuProT 1107

1571 GTCTAACCCGTGGT............GATGTACGGGTG......ATACAA 1602
     ::|||   :::|||            ||||||
1107 hrLeuProGlnGlyThrThrValThrAspValThrProGlyGlyThrIle 1123

1603 CAAACCTCGGCACCTGATAAACATGGGGTTTATCAAGCGACAGTGGAAAT 1652
     :::|||::::::|||         |||   |||:::::::::::::|||::
1124 AspThrAsnThrPro.........GlyAsnTyrGluGlyValIleGluVa 1137

1653 TAAAAAGCCTGATGGAAGTTGGGAGGTGAAAACGAAAAAAGGTGGGAAAG 1702
     :      |||||||||              ||||||         ||||
1137 lThrTyrProAspGly..............ThrLysAspThrValLysV 1149

1703 TGATGACCAAGCACACCATGTTCCCAAAAGATTGGGAT......GAGGCT 1746
     ||   ::::::   |||          :::|||   |||      |||
1149 alProValGluValThrAspAsnArgSerAspAlaAspLysTyrGluPro 1165

1747 AGAATTAGGGCTGAAGTTACTTCGGCTTGGGAAAGTAGAATAATGCTTAA 1796
       :::::::::|||   ::::::         :::::::::   |||
1166 ThrValGluGlyGluLysValGluIle...GlyGlyLysValAspLeuTh 1181

1797 GGATAATAAATGG.................CAGGGTACAAGTAAATCGG 1828
      ||||||                      |||||||||:::   :::
1181 rAspAsnValThrAsnLeuProThrLeuProGlnGlyThrThrValThrA 1198

1829 GTATTAAAATAGAAGGATTTACCGAACCTAAT 1860
       :::         |||      :::   |||
1198 spValThrProGlyGlyThrIleAspThrAsn 1208

seq_name: pir2:S24169

seq_documentation_block:
mucin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 10-Nov-1995
C;Accession: S24169
R;Huan, L.J.; Xu, G.; Forstner, G.; Forstner, J.
Biochim. Biophys. Acta 1132, 79-82, 1992
A;Title: A serine, threonine and proline-rich region near the carboxyl-terminus of a rat

A;Reference number: S24169; MUID:92379096
A;Accession: S24169
A;Molecule type: mRNA
A;Residues: 1-292 <HUA>


alignment_scores:
            Quality:  147.00              Length:    312
              Ratio:  1.185                 Gaps:     13
 Percent Similarity:  39.744   Percent Identity: 24.038

alignment_block:
 US-09-303-518D-463 x S24169   ..

 Align seg 1/1  to: S24169  from: 1  to: 292

 287 ATCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCC 336
     :::|||   |||:::   |||      ||||||:::   |||
  29 ValThrThrThrThrIleIleGlySerSerThrSerThrProThrThrTh 45

 337 GATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCAT 386
      |||       |||      |||   |||   |||      ::::::|
  45 rIleThrGluThrProSerProProSerThrIleSerGluThrAlaSerS 62

 387 CCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCAC 436
     ||   :::      |||::::::         |||:::   |||
  62 erThrSerProThrThrGluSerThrSerSerThrSerThrThr..... 76

 437 AGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTAC 486
           :::      |||   |||                    |||   ||
  77 ......SerGlnThrProThrPro.................ThrProTh 85

 487 GACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCG 536
     ||||   |||      |||      |||:::||||||   |||||||||:
  85 rThrIleLysGluThrPro...ThrSerThrSerThrValProThrThrG 101

 537 CAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGC 586
     :::::   ::::::      ||||||   |||:::   |||
 101 lySerThrSerSerLysProProThrGlySerSerThrProIle..... 115

 587 TGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCC 636
                                                 |||:::||
 116 .....................................ThrSerPr 118

 637 GAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGA 686
     ||||                ||||||      ||||||:::|||:::
 118 oSer.........ThrProSerProProThrGluSerThrThrLeuSerS 132

 687 TATCGTCAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCG 736
       :::      |||:::::::|||   ::::::      |||:::
 132 erThrProValThrThrThrAlaThrSerSerThrThrSerSer..... 146

 737 ATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGC 786

 146 ................................................. 146

 787 TTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGC 836
                    |||   |||
 147 ................ProGlyThr........................ 149

 837 AGATATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGG 886
                              |||   |||          |||::::::
 150 ....................ThrSerProPheAlaThrSerSerVal... 158

 887 CAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATC 936
     :::|||      |||   |||                |||||:::   ||
 159 SerSerThrProProSerPro..............ProSerSerThrSe 170

 937 TTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATA 986
     |         |||:::   |||:::|||:::      |||:::::::::|
 170 r......GlyProThrThrSerSerGlyMetProThrSerSerLysThrT 185

 987 CGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCG 1036
```

```
      ||:::
    185 hrThr..................................... 186

   1037 CGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGAT 1086
                           :::|||   |||   |||      |||
    187 .....................GlyProThrSerProThrThrArgPro.. 195

   1087 GCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTC 1136
                 |||:::|||          |||   |||   :::   ::::
    196 ..........ProSerThrSerThrProThrSerProThrValProThrS 209

   1137 AAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGC 1186
        ::|||   ::::::::::|||      :::   |||         |||
    209 erThrThrGluAlaIleThrGlnThrArgLeuSerSerThrThrPro... 224

   1187 CGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACC 1222
              :::|||::::::   |||:::|||   |||
    225 ThrMetGluThrThrArgThrSerSerTrpGlyThr 236
```

seq_name: pir2:A48018

seq_documentation_block:
mucin 7 precursor, salivary - human
N;Alternate names: mucin, MG2; mucin, MG2a-T1; mucin, MG2a-T2; mucin, MG2b-T2
C;Species: Homo sapiens (man)
C;Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 07-May-1999
C;Accession: A48018; S29115; S29116; S29114
R;Bobek, L.A.; Tsai, H.; Biesbrock, A.R.; Levine, M.J.
J. Biol. Chem. 268, 20563-20569, 1993
A;Title: Molecular cloning, sequence, and specificity of expression of the gene encoding
A;Reference number: A48018; MUID:93388636
A;Accession: A48018
A;Molecule type: mRNA
A;Residues: 1-377 <BOB>
A;Cross-references: GB:L13283
A;Experimental source: submandibular gland
A;Note: sequence extracted from NCBI backbone (NCBIN:137719, NCBIP:137720)
R;Reddy, M.S.; Bobek, L.A.; Haraszthy, G.G.; Biesbrock, A.R.; Levine, M.J.
Biochem. J. 287, 639-643, 1992
A;Title: Structural features of the low-molecular-mass human salivary mucin.
A;Reference number: S29114; MUID:93075006
A;Accession: S29115
A;Molecule type: mRNA
A;Residues: 143-168 <RED>
A;Accession: S29116
A;Molecule type: protein
A;Residues: 'S',71-79,'N',81-86,'XX',89,'X',91,'P' <RE2>
A;Accession: S29114
A;Molecule type: protein
A;Residues: 143-145,'X',147,'XXX',151-152,'X',154-158,'X',160-161,'A',163-164,'XX',167-1
C;Genetics:
A;Gene: GDB:MUC7
A;Cross-references: GDB:138799; OMIM:158375
A;Map position: 4q13-4q21
C;Keywords: glycoprotein
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-377/Product: mucin 7, salivary #status predicted <MAT>
F;97,128,135,146,312/Binding site: carbohydrate (Asn) (covalent) #status predicted

alignment_scores:
Quality: 147.00 Length: 383
Ratio: 0.987 Gaps: 23
Percent Similarity: 38.903 Percent Identity: 25.326

alignment_block:
US-09-303-518D-463 x A48018 ..

Align seg 1/1 to: A48018 from: 1 to: 377

```
     119 AGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGAG 168
         |||:::   |||||||||      :::|||   :::      :::   ::
      99 SerValValAsnProThrLeuValAlaThrThrGlnIleProSerValTh 115

     169 CTTGCCNAGCGCA...ACGGCCATATCGGATTGGGAAACATACAAAGCCA 215
         ::::|||||||||    |||   |||:::         |||        :
     115 rPheProSerAlaSerThrLysIleThr.........ThrLeuProAsnV 129

     216 TCAGTTGG............GCCACCTGATGATTCAACAGGCGGCCGTTG 253
         ::::::::            ||||||         :::|||:::   :::
     129 alThrPheLeuProGlnAsnAlaThrThrIleSerSerArgGluAsnVal 145

     254 AAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTC 303
         :::      |||::::::   :::      |||:::         |||||
     146 AsnThrSerSerSerValAlaThrLeuAlaProVal.........AsnSe 159

     304 CATTCGCCCT......TCGACAACCATGCCTCACATTCCGATTCTGACGA 347
         |      |||      :::|||:::   |||      |||      |||
     159 rProAlaProGlnAspThrThrAlaAlaProProThrProSerAlaThrT 176

     348 AGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACG 397
           |||:::   |||                  ::::::|||
     176 hrProAlaProPro.................SerSerSerAlaProPro 186

     398 GATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCG..... 442
         :::|||::::::   |||||||||      :::      :::|||
     187 GluThrThrAlaAlaProProThrProSerAlaThrThrGlnAlaProPr 203

     443 ..........................GCGGCTATCCCGCTCCCAAAGG 464
                                    ||||||   |||   |||
     203 oSerSerSerAlaProProGluThrThrAlaAlaProProThrPro.... 218

     465 CGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCC 514
                          |||||||||            |||      |||
     219 ..............ProAlaThrThr......ProAlaProProSerSer 228

     515 GCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGT 564
         ::::::          ||||||||||||||||            ||||||
     229 SerAlaProProGluThrThrAlaAlaPro............ProThrPr 241

     565 TTCC...ACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATT 611
          |||   |||   ||||||                              :::|
     241 oSerAlaThrThrProAla.....................ProLeuSerS 251

     612 CAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCG 661
         ||:::|||||||||:::|||:::                  |||:::|||
     251 erSerAlaProProGluThrThr.................AlaValPro 261

     662 CCGAAGCCTTCA...ACGGCACTGCAGATATCGTCAAAAACATCATCGGC 708
         |||   ||||||   |||:::|||:::   ||||||   ::::::
     262 ProThrProSerAlaThrThrLeuAspProSerSerAlaSerAla..... 276

     709 GCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGA 758

     276 ..................................................... 276

     759 AGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAA 808
                                                  ||||||:::
     277 .....................................ProProGlu 279

     809 ACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGAC 858
         |||                                          ::
     280 Thr.....................................ThrAl 282

     859 TATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGC 908
         :   |||      |||||||||                |||||   |||
     282 aAlaProProThrProSerAla.............ThrThrProAlaPro. 294

     909 ACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCA 958
                   ||||||                        ||||||:::
     295 ..............ProSer.....................SerProAla 299

     959 AAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCATCACGGCA 1008

     299 ..................................................... 299
```

```
    1009 CATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAA 1058

     299 ..................................................... 299

    1059 ATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAAATACCCGT 1108
            |||   :::::::|||      |||:::      ||||||||||:::
     300 ..ProGlnGluThrThrAlaAlaProIleThr...ThrProAsnSerSer 314

    1109 CCCCTTACCATT...CCCGAAATATCCGTT..................... 1135
         |||   |||:::   |||:::   |||
     315 ProThrThrLeuAlaProAspThrSerGluThrSerAlaAlaProThrHi 331

    1136 .CAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCG 1183
          ||||||   :::||||||||||:::::   |||:::      ||||||
     331 sGlnThrThrThrSerValThrThrGlnThrThrThrThrLysGlnPro 347

seq_name: pir1:S48478

seq_documentation_block:
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR019c
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C;Accession: S48478; A26877; B26877; S27281; JC6123
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN00009; MIP
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A;Reference number: A91831; MUID:87194600
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 1-242 <YAM>
A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A;Accession: B26877
A;Molecule type: DNA
A;Residues: 762-1331 <YA2>
A;Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A;Reference number: S27281; MUID:89031230
A;Accession: S27281
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A;Reference number: JC6123; MUID:96323237
A;Accession: JC6123
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1367 <LAM>
A;Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C;Genetics:
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A;Cross-references: MIPS:YIR019c; SGD:S0001458
A;Map position: 9R
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F;5-21/Domain: transmembrane #status predicted <TM1>
F;1350-1366/Domain: transmembrane #status predicted <TM2>


alignment_scores:
            Quality:  146.00            Length:    438
              Ratio:  0.737               Gaps:     16

 Percent Similarity:  45.205   Percent Identity: 23.288

alignment_block:
 US-09-303-518D-463 x S48478   ..

 Align seg 1/1  to: S48478  from: 1  to: 1367

      20 TATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGC......CGATGCAT 63
         :::|||   |||      ||||||:::      |||      :::|||
     192 PheProGlyPhe......TyrTrpAsnIleAspCysAspAsnAsnCysGl 206

      64 GCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGA 113
            |||                                  :::   ||||
     206 yGlyThr.............................LysSerSerT 212

     114 CCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGG 163
         ||:::::   |||:::      :::::||||||   |||:::::::::
     212 hrThrThrSerSerThrSerGluSerSerThrThrThrSerSerThrSer 228

     164 GGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAAGC 213
            |||      ::::::::::::   |||
     229 GluSerSerThrThrThrSerSerThrSer..................... 238

     214 CATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATAT 263
                                                  |||   |
     239 .....................................GluSerS 241

     264 CGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCT 313
         ||:::|||   |||:::         |||:::   :::|||
     241 erThrThrThrSerSer.........ThrSerGluSerSerThrSerSer 254

     314 TCGACAACCATGCCTCACATTCCGATTCTGACG..............AA 348
         ||||||||   |||      |||      |||               ||
     255 SerThrThrAlaProAlaThrProThrThrThrSerCysThrLysGluLy 271

     349 GCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGG 398
         ||||:::   |||   |||   |||:::                 |||
     271 sProThrProProThrThrThrSerCys.................ThrL 282

     399 ATACGAACACCATCCCGCCG............................. 418
            :::      ||||||
     282 ysGluLysProThrProProHisHisAspThrThrProCysThrLysLys 298

     419 .........ACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCC 459
                 |||:::   |||      :::   ::::::   |||:::||
     299 LysThrThrThrSerLysThrCysThrLysLysThrThrThrProValPr 315

     460 AAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAA 509
         |          :::      |||   ::::::         :::|||
     315 oThrProSerSerSerThrThrGluSerSerSerAlaProValProThrP 332

     510 TATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCG 559
           |||:::||||||      :::::::::||||||  ::::::
     332 roSerSerSerThrThrGluSerSerSerAlaProValThrSerSerThr 348

     560 ACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGA 609
         |||   |||:::         :::         ::::::   :::|||::
     349 ThrGluSerSerSerAlaProValProThrProSerSerSerThrThrGl 365

     610 TTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGC 659
         :|||::::::   |||   |||:::   |||   |||:::   :::
     365 uSerSerSerAlaProValThrSer...SerThrThrGluSerSerSerA 381

     660 CGCCGAAGCCTTCAACGGCACTGCAGATATCGTCA............AAA 697
           |||      |||::::::   :::   ||||||
     381 laProValThrSerSerThrThrGluSerSerSerAlaProValProThr 397

     698 ACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAG 747
            ||||||:::      |||:::   ||||||   :::
     398 ProSerSerSerThrThrGluSerSerSerAlaProValThrSerSerTh 414

     748 GGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCT 797
          :::   :::::::::      :::   |||   |||:::
```

```
  414 rThrGluSerSerSerAlaProValThrSerSerThrThrGluSerSerS 431

  798 TTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGC 847
          |||   |||          :::||||||
  431 erAlaProValThr.........SerSerThr..............Thr 439

  848 AACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAAC 897
      :::|||::::::   |||         |||:::      :::|||:::::
  440 GluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAl 456

  898 CCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATA.........TCTT 938
      :|||:::|||               ||||||:::            |||
  456 aProValPro............ThrProSerSerSerThrThrGluSerS 469

  939 TATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACG 988
              |||      |||:::      |||   |||:::      :::
  469 erSerAlaProValThrSerSerThrThrGluSerSerSerAlaProVal 485

  989 GCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCG 1038
              :::|||           ||||||:::         :::
  486 ProThrProSerSerSerThrThrGluSerSerSerAlaProValThrSe 502

 1039 ATCGCATTGCCGAAAGGGAAATCCGCCG.................TCAG 1070
       |||      ::::::::::::   |||                  |||:
  502 rSerThrThrGluSerSerSerAlaProValProThrProSerSerSerT 519

 1071 CGACAATTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATT 1120
      ::|||                 :::|||:::              |||
  519 hrThrGluSerSerSerAlaProAlaProThrProSerSerSerThrThr 535

 1121 CCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATC 1170
          :::   |||:::   :::   ||||||:::|||   ::::::::::
  536 GluSerSerSerAlaProValThrSerSerThrThrGluSerSerSerAl 552

 1171 ACCT...CCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGC 1217
      :|||   |||   |||            |||   :::   |||:::
  552 aProValProThrProSerSerSerThrThrGluSerSerSerThrProV 569

 1218 AGACCAACGCCACC 1231
        |||::::::|||
  569 alThrSerSerThr 573
```

seq_name: pir2:I47141

seq_documentation_block:
gastric mucin (clone PGM-2A) - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 03-Nov-2000
C;Accession: I47141; S55315
R;Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
Gastroenterology 106, 200, 1994
A;Title: Pig gastric mucin: isolation and characterization of a cDNA clone with a novel
A;Reference number: I47141; MUID:94102478
A;Accession: I47141
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-528 <TUR>
A;Cross-references: EMBL:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208
R;Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
Biochem. J. 308, 89-96, 1995
A;Title: Isolation and characterization of cDNA clones encoding pig gastric mucin.
A;Reference number: S55315; MUID:95275264
A;Accession: S55315
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-528 <TU2>
A;Cross-references: GB:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208
C;Superfamily: pig submaxillary mucin

alignment_scores:
Quality: 145.50 Length: 479
Ratio: 0.591 Gaps: 21

Percent Similarity: 51.357 Percent Identity: 23.173

alignment_block:
US-09-303-518D-463 x I47141 ..

Align seg 1/1 to: I47141 from: 1 to: 528

```
   92 CCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAA 141
      |||:::|||   :::   |||:::   |||   |||:::||||||   ::
    1 ProIleSerValGlnProSerSer...SerSerSerSerProThrThrSe 16

  142 TACCACC............TATTCGGCAGCAGGGGGGAGCTTGCCNAGCG 179
      :||||||               |||:::::::::      :::|||||||:
   16 rThrThrSerValGlnSerSerSerSerSerSerValProIleProSerT 33

  180 CAACGGCCATA..............TCGGATTGGGAAACATACAAAGCC 214
      ::|||::::::               ::::::      |||   :::|||
   33 hrThrSerValGlnProSerSerSerGlySerAlaProThrThrSerAla 49

  215 ATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATC 264
         |||      |||         :::      |||        |||||
   50 ThrSerValGlnThrSerSerSerSerSerProPro.........IleSe 63

  265 GGCTACATTG.....................TCCGCTTTTCCGATCACGG 293
      |:::|||:::                     |||:::   |||   |||:
   63 rSerThrIleSerValGlnThrSerSerSerSerSerValProThrThrS 80

  294 GCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTG 343
      ::|||:::|||:::::::||||||::::::
   80 erThrThrSerValGlnProSerSerSerSerSer......AlaProThr 94

  344 ACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGG 393
      |||:::   :::         ::::::|||:::   :::|||   |||::
   95 ThrArgAlaThrSerValGlnSerSerSerSerSerSerAlaProIleSe 111

  394 GACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGG 443
      ::::   |||::::::   |||   :::                  ::::
  111 rSerThrThrSerValGlnProSerSer.................SerG 122

  444 CGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAA 493
      :::::::::|||      :::|||   ::::::   ::::::::::::
  122 lySerValProThrThrSerAlaThrSerValGlnSerSerSerSerSer 138

  494 AAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACC 543
      :::|||            ::::::|||:::   |||:::::::::::::
  139 SerAlaProThrThrSerAlaThrSerValGlnProSerSerSerSerSe 155

  544 GGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCA 593
                  |||||||||:::::|||
  155 rProProIleSerSerThrValSerValGlnPro.............. 166

  594 AGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGG 643
             :::::::::|||:::|||||||      :::|||   |||
  167 ........SerSerSerSerSerAlaProThrThrSerAlaThrSerVal 180

  644 ......ACAGATCGGGCAATGCCGCCGAAGCCTTCAACG...GCACTGCA 684
            ::::::   :::   ||||||      ||||||   :::::::||
  181 GlnProSerSerSerSerSerProProIleSerSerThrValSerValGl 197

  685 GATATCGTCAAAA..............ACATCATCGGCGCGGCAGGAGA 719
      |   ||||||:::               ||||||::::::      :::
  197 nThrSerSerSerSerSerValProThrThrSerThrThrSerValGlnP 214

  720 AATTGTCGGCGCAGGCGATGCCG.................TGCAGGGTA 751
            |||:::::::::::::|||                    |||
  214 roSerSerSerSerSerValProThrThrSerAlaThrSerValArgSer 230

  752 TAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCC 801
          ::::::::::::|||   :::      |||:::   |||
  231 SerSerSerSerSerThrProIleProSerThrThrSerVal......Gl 245

  802 ACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACT 851
       |||::::::         |||   |||                    :
```

```
    245 nProSerSerSerSerSerAlaProThr.....................T 255

    852 CAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCA 901
        ::::::::         ||||||||||:::      :::|||:::||||||
    255 hrSerAlaThrSerValGlnProSerSer......SerSerSerThrPro 269

    902 ATGCCG.............................CACAAGGCATAGAA 921
        :::|||                                  :::|||
    270 IleProSerThrThrSerValGlnProSerSerSerSerSerAlaProTh 286

    922 GCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGC 971
            ||||||   |||:::   |||||||||   |||::::::      :
    286 rThrSerAlaThrSerVal...GlnProSerSerSerSerSerProProI 302

    972 TG.................................TCCGGGGAAAATACG 988
        ::                                  |||:::::   |||
    302 leSerSerThrIleSerValGlnProSerSerSerSerSerSerProThr 318

    989 GCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCG 1038
        :::   ::::::|||   :::      |||||||||          :::
    319 ThrSerThrThrSerValGlnProSerSerSerGlySerAlaProThrTh 335

   1039 ATCG..............CATTGCCGAAAGGGAAATCCGCCGTCAGCGA 1073
         |||                        :::::::::||||||   ::::
    335 rSerAlaThrSerValGlnProSerSerSerSerSerProProIleSerS 352

   1074 CAATTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCC 1123
        ::   :::   ::::::   :::   ::::::        |||
    352 erThrIleSerValGlnProSerSerSerSerSerSerProThrThrSer 368

   1124 GAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACC 1172
             |||||||||      ||||||   :::|||      ||||||
    369 ThrThrSerValGlnProSerSerSerGlySerAlaProThrThrSerAl 385

   1173 ......................TCCTCAACCGTGCCGCCGTCAAACGGCA 1201
                               ||||||:::||||||   :::::::::
    385 aThrSerValGlnProSerSerSerSerSerValProThrThrSerAlaT 402

   1202 AAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTT 1251
          :::|||:::   :::::::::         :::|||   :::|||
    402 hrSerValArgSerSerSerSer......SerSerThrProIleProThr 416

   1252 GACGGTAAAGGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAA 1301
              :::      |||:::   ::::::   |||        :::
    417 ThrThrSerValGlnProSerSerSerSerSerValProThrThrSerAl 433

   1302 GCTCGATATTCAAGAATTATCGGGGGGCGGTATACCT 1338
            ::::::|||      |||:::::::::   |||
    433 aThrSerValGlnThrSerSerSerSerSerThrPro 445
```

seq_name: pir2:T14180

seq_documentation_block:
exiT protein - Mycobacterium smegmatis
C;Species: Mycobacterium smegmatis
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 01-Dec-2000
C;Accession: T14180; T14163
R;Zhu, W.M.; Arceneaux, J.E.L.; Beggs, M.L.; Byers, B.R.; Eisenach, K.D.; Lundrigan, M.D
Mol. Microbiol. 29, 629-639, 1998
A;Title: Exochelin genes in Mycobacterium smegmatis: identification of an ABC transporte
A;Reference number: Z17906; MUID:98385832
A;Accession: T14180
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1122 <ZHU>
A;Cross-references: EMBL:AF034152; NID:g3421057; PID:g3421058; PIDN:AAC32046.1
R;Yu, S.; Fiss, E.; Jacobs Jr., W.R.
J. Bacteriol. 180, 4676-4685, 1998
A;Title: Analysis of the exochelin locus in mycobacterium smegmatis: biosynthesis genes
A;Reference number: Z17898; MUID:98389687
A;Accession: T14163
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 534-1122 <YUS>
A;Cross-references: EMBL:AF027770; NID:g3560502; PID:g3560505; PIDN:AAC82548.1
C;Genetics:
A;Gene: exiT
C;Superfamily: Mycobacterium tuberculosis probable ABC transporter Rv0194; ATP-bindin

alignment_scores:
Quality: 144.00 Length: 385
Ratio: 0.966 Gaps: 23
Percent Similarity: 38.701 Percent Identity: 27.013

alignment_block:
US-09-303-518D-463 x T14180 ..

Align seg 1/1 to: T14180 from: 1 to: 1122

```
     69 CGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCGTC 118
        |||   ::::::|||:::|||:::      |||:::   ::::::|||
    261 ArgAlaGluLeuGlySerArgGly......ProGlyValGlyGlnProAr 275

    119 AGCATTTCGAACCCGACGGGAAATACCACCTATT...CGGCAGCAGGGGG 165
              |||   :::|||      |||||||::   |||   :::||||
    275 gArgSerArgProGluArgProAspProProLeuProArgLeuArgGlyG 292

    166 GAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAAGCCA 215
        ||                  |||   |||   |||            |||
    292 lyPro..............ProArgArgArgGly............Pro 299

    216 TCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATCG 265
           |||||||||      ::::::   |||:::
    300 ProValGlyProGlnArgAsnThrArgGlySer................. 310

    266 GCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTC 315
           :::   |||:::   |||   ||||||:::
    311 ....GluLysProValValArgArgArgAlaArg..............  320

    316 GACAACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGA 365

    320 .................................................. 320

    366 CGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCG 415
        ||||||:::          |||                |||   |||
    321 ArgIleArgArgGlyThrHisGlu............ArgArgProArgCy 333

    416 CCGACGGCTATGACGGGCCACAGGGCGG...CGGCTATCCCGCTCCCAAA 462
         ||||||      |||::::::   |||         ||||||:::||||
    333 sArgArgThrHisArgSerAlaLeuArgIleSerAlaSerArgGlyGlnA 350

    463 GGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATAT 512
        ||:::   :::                        |||   |||
    350 rgHisGlyAla.....................GlyArgTrpProAlaPro 359

    513 CCGCCTCAACCT...GACCGACAACCGCAGCACCGGACAACGGCTTGCCG 559
           |||:::|||   :::|||   |||      ||||||   |||   ||
    360 ArgProAspProAlaGluArgAlaPro......ArgThrProAlaArgAr 374

    560 ACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGT.......... 599
        |      ||||||               ||||||   :::
    374 gThrProProGlnLeuValValGlnArgAspAlaGlyAsnGlnSerGluC 391

    600 .....AGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGA 644
             |||:::|||      :::|||      ::::::   |||
    391 ysHisArgAsnArgSerAlaAlaArgThrGlyLeuArg...ArgArgArg 406

    645 CAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCA 694
           |||      ||||||      :::|||   :::      |||::
    407 ThrThrGlyValValArgArgArgArgArgArgProAlaProArgArgSe 423

    695 AAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGA....... 737
        ::::      |||||||||   |||            ||||||
    423 rArgProCysArgArgGlyAlaArgCysGluProValArgArgThrThrP 440
```

```
 738 ..................TGCCGTGCAGGGTATAAGCGAAGGCTCAAA 767
                        |||   ||||||
 440 roThrLeuAspProGlyProCysAlaAlaGly.................. 450

 768 CATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGG 817

 450 ..................................................... 450

 818 CGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCA 867
           |||   ||||||   ||||||           :::||||
 451 ..........ArgSerGlyArgProGlyAla............AlaArgS 460

 868 GCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCAT 917
     ||:::|||                              |||   ||||||
 460 erAspHis............................ArgAspArgHis 466

 918 AGAAGCCGT.........................CAGCAATATCTTTA 940
     |||   |||                        :::
 467 ArgHisArgAlaGlyHisArgArgArgHisThrProAlaAlaHisAlaAr 483

 941 TGGCAGCCAT.........CCCCATCAAAGGGATTGGAGCTGTCCGGGGA 981
      |||:::|||        |||||||||::::::           |||
 483 gGlyThrHisAsnGlyGlyProHisGlnGlnSer............GlyP 496

 982 AAATACGGCTTGGGCGGCA.......TCACGGCACATCCTGTCAAGCGGT 1024
          |||   ||||||         :::||||||      :::||||||
 496 roAlaArgArgGlyArgProGlyAlaAlaArgHisArgArgAlaSerGly 512

1025 CGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGAC 1074
     :::|||
 513 GlnArg..................................................... 514

1075 AATTTTGCCGATGCGGCATACGCCAAATACC......CGTCCCCTTACCA 1118
                |||||||||||:::::      |||||||||   :
 515 .............ArgHisThrProArgSerGlyArgArgProLeuArgL 527

1119 TTCCC 1123
     ::|||
 527 euPro 528

seq_name: pir2:B47328

seq_documentation_block:
natural killer cell tumor-recognition protein - mouse
N;Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumo
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 05-Nov-1999
C;Accession: B47328; I77662
R;Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A;Title: A cyclophilin-related protein involved in the function of natural killer cells.
A;Reference number: A47328; MUID:93133824
A;Accession: B47328
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1507 <AND>
A;Cross-references: GB:L04289; NID:g192866
A;Note: authors translated the codon AGT for residue 972 as Arg
R;Rinfret, A.; Anderson, S.K.
Mol. Immunol. 30, 1307-1313, 1993
A;Title: IL-2 regulates the expression of the NK-TR gene via an alternate RNA splicing m
A;Reference number: I57820; MUID:94019422
A;Accession: I77662
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 235-237;263-294 <RIN>
A;Cross-references: GB:S65998; NID:g425701; PIDN:AAB28500.1; PID:g425702
C;Genetics:
A;Gene: NK-TR
C;Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C;Keywords: alternative splicing; lymphocyte
F;60-230/Domain: cyclophilin homology <CYP>
```

```
alignment_scores:
           Quality:  143.50              Length:    507
             Ratio:  0.595                 Gaps:     29
 Percent Similarity:  47.535   Percent Identity: 25.049

alignment_block:
 US-09-303-518D-463 x B47328   ..

 Align seg 1/1  to: B47328  from: 1  to: 1507

 304 CATTCGCCCTTCGACAACCAT......GCCTCACATTCCGATTCTGACGA 347
     ||||||      |||::::::         |||||||||   ||||||::
 567 HisSerSerArgAspSerTyrArgSerLysSerHisSerArgSerAspSe 583

 348 AGCCGGTAG.TCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGAC 396
     :   |||   ||||||   |||
 583 rArgGlySerSerArgSerArg.........................A 591

 397 GGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGG 446
       :::         |||   |||   |||   |||:::   :::|||
 591 laValSerLysSerSerSerArgSerLeuAsnArgSerLysSerArgSer 607

 447 CTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAG 496
        |||||||||      |||:::      |||:::
 608 SerSerArgSerGlyProArgArgThrSerIleSer.............. 619

 497 GCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGA 546
           ||||||   |||   |||   :::::::::          :::|
 620 .......ProLysLysProAlaGlnLeuSerGluAsnLysProValLysT 634

 547 CAACGGCTTGCCGACCG...TTTCCACAATGCCGGCGCTATGCTGACGCA 593
     ||          ||||||      ||||||    |||   :::|||:::||
 634 hrGluProLeuArgProSerValProGln.AsnGlyAsnValLeuValGl 650

 594 AGGAGTAGGCGACGGAT.....................TCAAACGCGCCA 622
     |   ||| |||::::::                      |||::::::|||
 650 nProVal.AlaAlaGluAsnIleProValIleProLeuSerAspSerPro 666

 623 CCCGATACAGCCCCGAGC...TGGACAGATCGGGCAATGCCGCCGAAGCC 669
     |||          ||||||   |||                |||   |||||
 667 Pro.........ProSerArgTrpLysProGlyGlnLysProTrpLysPr 680

 670 TTCA......ACGGCACTGCAGATATCGTCAAAAACA.............. 700
     ||||            ::::::::::::||||||
 680 oSerTyrGluArgIleGlnGluMetLysAlaLysThrThrHisLeuLeuP 697

 701 ........TCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCG 742
             |||:::            :::::::::|||   :::
 697 roValGlnSerThrTyrSerLeuThrAsnIleLysAlaThrVal.SerSe 713

 743 TGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGT 792
      ::::::|||   :::|||   |||                    ||||
 713 rSerSerTyrHisLysArgGluLysPro......SerGluSerAspGlyS 728

 793 CTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATAT 842
     |||||:::   :::   :::|||   :::              ||||||
 728 erAlaTyrSerLysTyrSerAspArgSerSerGlySerSerGlyArgSer 744

 843 GGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCC 892
     |||:::   :::          |||||||||
 745 GlySerLysSerSerArgSerArgSerSer..................... 754

 893 AAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATG 942
                 |||:::      |||||||||::::::   |||
 755 .........SerArgSerTyrThrArgSerArgSerArgSerLeuProT 768

 943 GCAGCCAT..............CCCCATCAAAGGGATTGGAGCTGTCCG 977
       |||:::              |||   :::|||:::   |||   |||
 768 hrSerArgSerLeuSerArgSerProSerSerArgSerHisSer...Pro 783

 978 GGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGC 1027
     :::|||         |||   :::|||:::   |||   :::   :::::
```

```
 784 AsnLysTyrSerAspGlySerGlnHisSerArgSerSerSerTyrThrSe 800

1028 A............GATGGGCGCGATCGCATT......GCCGAAAGGGAAA 1059
     :           |||||||||      :::      ::::::||| |||
 800 rValSerSerAspAspGlyArgArgAlaMetPheArgSerAsnArg.Lys 816

1060 TCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAAATACCCGTC 1109
     ::::::|||:::::::::            :::   ::::::      ::
 817 LysSerValThrSerHisLysArgHisArgSerAsnSerGlu......Ly 831

1110 CCCTTACCATTCCCGAAATATCCGTTCA...AACTTGGAGCAGCGTTACG 1156
     :      ||||||:::   :::|||:::   :::   ::::::|||:::
 831 sThrLeuHisSerLysTyrValArgGlyArgGluLysSerSerArgHisA 848

1157 GCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAAT 1206
       |||   :::   :::   |||::::::      ::::::
 848 rgLysTyrSerGluSerArgSerSerLeuAspTyrThrSerAsp..... 862

1207 GTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGG 1256
             :::||||||   |||   ::::::
 863 .........SerAspGlnSerHisValGlnVal................. 870

1257 TAAAGGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCG 1306
        :::   |||:::   ||||||:::   |||   ::::::      |
 871 .TyrSerAlaProGluLysGluLysGlnGlyLysValGluAlaLeuAsnA 887

1307 ATATTCAAGAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTT 1356
     ||   |||          |||   |||       ::::::||||||
 887 spLysGln.........GlyLysGlyArgGluGluGlyLysPro...... 898

1357 GATGCGAAACCGAGATGGGAGGTTGATAGGAAGCTTAATAAATTGACAAC 1406
           ||||||:::|||||||      |||
 899 ......LysProGluTrpGluCysProArg..................... 906

1407 TCGTGAGCAGGTGGAGAAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGA 1456
           :::   ::::::|||   ::::::      |||   ::::::
 907 .......SerLysLysGluAsnSerGluAspHisSerArgAspAspSerV 921

1457 GTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGGGAAAATAAAACAGGG 1506
     |||:::   |||      |||      :::|||:::::::::::::
 921 alSerLysGlyLysAsnCysAlaGlySerLysTrpAspSerGluSer... 936

1507 TTAGATTTTAATCATTTTATAGGTGGTGATATCAATAAGAAAGGCACAGT 1556
                                  |||::::::      ||
 937 ................................AsnSerGluGlnAspVa 942

1557 AACAGGAGGGCATAGTCTA......ACCCGTGGTGATGTACGGGTGATAC 1600
     ||||   ::::::::            ||||||:::   :::
 942 lThrLysSerArgLysSerAspProArgArgGlySerGluLysGluGluG 959

1601 AACAAACCTCGGCACCTGATAAACATGGGGTTTATCAAGCGACAGTGGAA 1650
       ::::::|||:::   ::::::      |||   |||:::   ::::::
 959 lyGluAlaSerSerAspSerGluSerGluValGlyGlnSerHisIleLys 975

1651 ATTAAAAAGCCT.................................GA 1664
        |||   |||                                 ||
 976 AlaLysProProAlaLysProProThrSerThrPheLeuProGlySerAs 992

1665 TGGAAGTTGGGAG 1677
     ||||:::|||:::
 992 pGlyAlaTrpLys 996
```

seq_name: pir2:A53577

seq_documentation_block:
ascites sialoglycoprotein 1 - rat (fragments)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 12-Apr-1995 #sequence_revision 12-Apr-1995 #text_change 07-Feb-1997
C;Accession: A53577
R;Wu, K.; Fregien, N.; Carraway, K.L.
J. Biol. Chem. 269, 11950-11955, 1994
A;Title: Molecular cloning and sequencing of the mucin subunit of a heterodimeric, bifun
A;Reference number: A53577; MUID:94216302
A;Accession: A53577
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1630 <WUA>
A;Cross-references: GB:U06746
C;Keywords: glycoprotein

alignment_scores:
Quality: 143.50 Length: 469
Ratio: 0.649 Gaps: 21
Percent Similarity: 47.122 Percent Identity: 22.388

alignment_block:
US-09-303-518D-463 x A53577 ..

Align seg 1/1 to: A53577 from: 1 to: 1630

```
  65 CACACGCCTCAGATTTGGCAAACGATCCCT.............TTATC 99
     |||:::|||:::      ::::::   |||             |||::
 720 HisAlaProArgGluSerSerSerProProSerThrSerAspMetLeuTh 736

 100 CGGCAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCT 149
     :         ||||||   :::   |||:::   |||||||:::|||||||
 736 rThrThrAlaSerThrGluGlyThrSerGlyAspThrGlyHisThrThrA 753

 150 ATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGG 199
       ::::::   ||||||:::|||:::::::|||   :::|||
 753 laValThrThrGlnGlySerIleProAlaThrThrGlnLeuSerThrThr 769

 200 .........GAAACATACAAAGCCATCAGTTGGGCCACCTGATGATTCAA 240
             :::      :::::::::|||      |||
 770 PheAlaSerGlnLysMetSerThrValSerThrProThr........... 782

 241 CAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCA 290
               ::::::|||:::|||   ||||||                |
 783 .ThrSerSerIleGlnGluLeuSer...ThrLeuProGlnSerGlnHisT 798

 291 CGGGCACAAATTCCATT......CGCCCTTCGACAACC............ 322
     |||||:::   :::|||      ||||||:::|||||
 798 hrGlySerMetGluIleSerSerArgProGlnThrThrSerValThrSer 814

 323 .................ATGCCTCACATTCCGATTCTGACGAAGCCGGT 354
                                  |||:::   |||:::   ||
 815 ThrLeuSerSerSerProSerGlySerThrProValGlnThrArgSerVa 831

 355 AGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGA 404
     |:::      :::|||:::      :::   :::   :::      :::|
 831 lThrSerSerSerAspGluArgThrAsnProThrSerSerGlyValSerA 848

 405 ACACCATCCCG.................................CCGACG 421
     |||||   |||                                  |||
 848 snThrSerProAlaThrThrGluValLeuThrProThrSerSerProGlu 864

 422 GCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGG 471
     :::      |||:::   |||                |||::::::
 865 SerThrProGlyAsnThrAla..............ProArgThrThrGl 876

 472 GATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAA 521
           |||:::||||||   |||:::|||                :::
 876 uThrSerThrThrThrThrThrLysValLeuMetThrSerLeuGlnGlnL 893

 522 CCTGACCGACA..............ACCGCAGCACCGGACAACGGCTTG 556
           ||||||              |||::::::|||   :::
 893 ysLeuProThrGlySerThrLeuGlyThrSerThrProThrGluValThr 909

 557 CCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGAC 606
        |||:::|||:::      :::          :::   |||   :::::
 910 ThrThrLeuSerAlaSerSerSerAspGlnValGlnValGluThrThrSe 926

 607 GGATTCAAACGCGCCACCCGAT.........ACAGCCCCGAGCTGGACAG 647
     ::::::::::      ||||||          |||:::   :::      :
 926 rGlnThrThrLeuSerProAspAlaThrThrThrSerHisAlaProArgG 943
```

```
     648 ATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAA 697
         ::   :::   ||||||:::   |||:::   |||         ::::::
     943 luSerSerSerProProSerThrSerValIleLeuThrThrMetAlaSer 959

     698 ACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAG 747
         |||::::::::::      :::         :::   |||:::      ::
     960 ThrGluGlyThrSerGlyAspThrGlyHisThrThrAlaValThrAspGl 976

     748 GGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCT 797
         :      :::   |||   |||   :::|||   |||
     976 nGlySerThrProAlaThrThrGluIleSerValThr..............  988

     798 TTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGC 847
              |||:::|||:::   :::   ||||||:::      :::
     989 .....ProSerThrGln...LysMetSerThrValSerThrLeuValThr 1002

     848 AACTCAAAG............ACTATGCCGCAGCAGCCATCCGCGATTGG 885
         :::::::::            ::::::   |||:::   ::::::::::||
    1003 SerThrGlnGluLeuThrSerSerGlnSerGlnArgThrGlySerMetGl 1019

     886 GCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATAT 935
         |   |||:::   |||         :::|||      |||:::   ::::
    1019 yThrSerSerLysPro.........GlnAlaThrThrProThrGluValT 1033

     936 CTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAAT 985
         ::                                              :::
    1033 hr.....................................ThrSer 1035

     986 ACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGC 1035
         |||      :::   ||||||      |||   |||           ||
    1036 ThrLeuSerSerPheSerArgGlySerLeuPheSer............Al 1048

    1036 GCGATCGCATTGC.........CGAAAGGGAAATCCGCCGTCAGCGACAA 1076
         ||||:::   |||         :::|||         |||:::::::::
    1048 aArgAsnCysCysLeuGlnThrLysLysProProLeuProAlaValValC 1065

    1077 TTTTGCCG.....................ATGCGGCATACGCCAAATACC 1105
           ||||||                     :::   |||:::   :::
    1065 ysLeuProAspProSerSerValProSerLeuMetHisSerSerLysPro 1081

    1106 CGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTAC 1155
         :::      |||      |||:::::::::::::|||   ||||||   ::
    1082 GlnAlaThrThrProThrGluValThrThrSerThrLeuSerSerPheSe 1098

    1156 GGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAA 1205
         :      :::|||                        ||||||      :
    1098 rArgGlySerThr.....................GlnThrGlnThrV 1107

    1206 TGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACG 1255
         ::|||   |||:::|||:::::::::   :::   :::         |||
    1107 alSer...TrpGluThrSerSerSerGlyLysIleThrAlaProSerThr 1122

    1256 GTAAAGGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTC 1305
            :::              |||   |||   ::::::      |||::
    1123 SerSer.............ArgArgThrProSerValAlaThrSerAs 1134

    1306 GATATTC 1312
         :||||||
    1134 pIlePhe 1136

seq_name: pir2:T38459

seq_documentation_block:
hypothetical divergent repeat-containing protein - fission yeast (Schizosaccharomyces pc
C;Species: Schizosaccharomyces pombe
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C;Accession: T38459; T38380
R;Harris, D.; McDonald, S.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z21794
A;Accession: T38459
A;Molecule type: DNA
A;Residues: 1-1748 <HAR>
A;Cross-references: EMBL:Z69368; PIDN:CAA93290.1; GSPDB:GN00066; SPDB:SPAC27F1.01c
R;McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21790
A;Accession: T38380
A;Molecule type: DNA
A;Residues: 1457-1794 <MCL>
A;Cross-references: EMBL:Z70691; PIDN:CAA94638.1; GSPDB:GN00066; SPDB:SPAC25G10.09c
A;Experimental source: clone c25G10
C;Genetics:
A;Gene: SPAC27F1.01c; SPDB:SPAC25G10.09c
A;Map position: 1


alignment_scores:
            Quality:  143.50              Length:    420
              Ratio:  0.829                Gaps:     20
 Percent Similarity:  41.190   Percent Identity: 22.143

alignment_block:
 US-09-303-518D-463 x T38459   ..

 Align seg 1/1  to: T38459  from: 1  to: 1794

      12 CCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATGC 61
         |||::::::          |||   :::|||:::   :::|||   :::||
    1326 ProSerThrThrSerThrSerPheAsnThrAlaProIleProGlnGlnAl 1342

      62 ATGCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTC 111
         |      :::   :::|||:::|||   ||||||   |||
    1342 aProLeuGluAsnGlnPheSerLysMetSerLeuGluPro.......... 1355

     112 GACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAG 161

    1355 ................................................. 1355

     162 GGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAA 211
                                  |||   |||   ::      |||::
    1356 .......................ProValArgProAlaValProThrSe 1364

     212 GCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAAT 261
         :|||:::         |||||||:::|||
    1364 rProLysProGlnIleProAspSerSer..................... 1373

     262 ATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCC 311
           :::::::|||:::|||     |||                 ::::::|
    1374 ..AsnValHisAlaProPro..Pro..............ProValGlnP 1384

     312 CTTCGACAACCATGCCT...CACATTCCGATTCTGACGAAGCCGGTAGTC 358
         ||   ::::::||||||   |||      :::   :::::|||   :::
    1384 roMetAsnAlaMetProSerHisAsnAlaValAsnAlaArgProSerAla 1400

     359 CCGTTGACGGATTCAGCCTTT.......................ACCGC 384
         |||         :::::::|||                      :::::
    1401 ProGluArgArgAspSerPheGlySerValSerSerGlySerAsnValSe 1417

     385 ATCCATTGGGACGGATACGAACACCATCCCG................... 415
         :||||||      :::|||:::|||:::|||
    1417 rSerIleGluAspGluThrSerThrMetProLeuLysAlaSerGlnProT 1434

     416 .....CCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCG...... 454
              |||   |||   ::::::|||         ::::::|||
    1434 hrAsnProGlyAlaProSerAsnHisAlaProGlnValValProProAla 1450

     455 .......................CTCCCAAAGGCGCGAGGGATATATAC 480
                                 |||||||||   |||:::   ||
    1451 ProMetHisAlaValAlaProValGlnProLysAlaProGlyMetValTh 1467

     481 AGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGA 530
         |   :::                  :::   |||:::   |||:
    1467 rAsnAlaPro.....................AlaProSerSerAlaProA 1477
```

```
 531 CAACCGCAGCACCGGACAACGGCTTGCCGACCGTT..............  565
     ::      ||||||   :::   ||||||   :::
1477 laProProAlaProValSerGlnLeuProProAlaValProAsnValPro 1493

 566 ......TCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGA 609
           |||   :::|||:::::::::   :::
1494 ValProSerMetIleProSerValAlaGlnGln.....ProPr 1506

 610 TTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGC 659
      |||::::::   |||   ||||||||||||   |||         :::|
1506 oSerSerValAlaProAlaThrAlaProSerSerThr.........LeuP 1520

 660 CGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCG 709
     |||||:::   |||:::      ::::::   |||   :::
1520 roProSerGlnSerSerPheAlaHisValProSerProAla..... 1533

 710 CGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAA 759

1533 ................................................. 1533

 760 GGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAA 809
                                          ||||||
1534 ....................................ProPro.... 1535

 810 CAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACT 859

1535 ................................................. 1535

 860 ATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCA........ 901
        ||||||:::|||||||||   :::   |||:::::|||
1536 AlaProGlnHisProSerAlaAlaAlaLeuSerSerAlaProAlaAspAs 1552

 902 ....ATGCCGCACAAGGCA.........................TAGA 920
         |||||||||::::::                          |
1552 nSerMetProHisArgSerSerProTyrAlaProGlnGluProValGlnL 1569

 921 AGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAG 970
     |||||:::||||||:::        |||:::   :::   |||   :::
1569 ysProGlnAlaIleAsnAsnIleAlaProAlaThrAsnLeuGlyThrSer 1585

 971 CTGTCCGGGGAAAATACGGCTTGGG......CGGCATCACGGCACATCCT 1014
        |||   :::    |||:::|||      ::::::   |||::: |||
1586 GlnSerPheSerPro.ArgMetGlyProValAsnAsnSerGlySer.Pro 1601

1015 GTCAAGCGGTCGCAGATGGGC......GCGATCGCATTGCCGAAAGGGAA 1058
     :::      :::      |||      ::::::|||:::|||
1602 LeuAlaMetAsnAlaAlaGlyGlnProSerLeuAlaValProAlaValPr 1618

1059 ATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAAATACCCGT 1108
      ||||||   |||::::::|||        |||      :::   |||:
1618 oSerAlaProSerAsnHisPheAsnProPheAlaLysMetGlnProProA 1635

1109 CCCCT 1113
     ::|||
1635 laPro 1636
```

seq_name: pir2:T29634

seq_documentation_block:
hypothetical protein C12D12.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T29634
R;Nhan, M.; Hawkins, J.
submitted to the EMBL Data Library, March 1996
A;Description: The sequence of C. elegans cosmid C12D12.
A;Reference number: Z20656
A;Accession: T29634
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-825 <NHA>
A;Cross-references: EMBL:U51998; PIDN:AAA96080.1; GSPDB:GN00028; CESP:C12D12.1
A;Experimental source: strain Bristol N2; clone C12D12

C;Genetics:
A;Gene: CESP:C12D12.1
A;Map position: X
A;Introns: 48/1; 86/3; 137/1; 172/3; 224/3; 253/1; 287/3; 328/2; 454/1; 487/3; 692/1
C;Superfamily: Epstein-Barr virus membrane antigen gp350

alignment_scores:
Quality: 143.00 Length: 376
Ratio: 0.851 Gaps: 16
Percent Similarity: 44.681 Percent Identity: 22.074

alignment_block:
US-09-303-518D-463 x T29634 ..

Align seg 1/1 to: T29634 from: 1 to: 825

```
 110 TCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGC 159
     :::|||||||||      :::|||      :::::::|||      :::::
 473 ThrThrValSerThrMetSerProProThrThrValThrValProThrTh 489

 160 AGGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACA 209
     :       :::|||::::::|||
 489 rProThrProValProThrThrThr........................ 497

 210 AAGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAA 259
                                     |||      |||   :::
 498 ............................AsnThrProProAlaAsnPro 504

 260 ATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCG 309
       :::||||||             |||   |||      :::|||   ::
 505 ThrThrAlaThr............ProThrThrValGlyThrSerLysGl 517

 310 CCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCC 359
     :   :::|||      |||||||::      :::|||      :::|||
 517 nThrAsnThrIleSerProHisLeuSerThrIleThrGlySerIleVal. 533

 360 CGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACC 409
                   :::   |||   :::::::::   :::|||::::::
 534 ..............ThrSerThrProThrMetAlaProGlnThrSerAla 545

 410 ATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCC 459
       |||   |||:::   |||   |||   :::|||:::   |||
 546 SerProThrThrThrProThr...HisThrThrAlaSerGlnProThrTh 561

 460 AAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAA 509
                           ::::::||||||   ::::::   |||:::
 561 rThrLysPro.........ValValThrThrAsnSerValThrProSerT 575

 510 TATCCGCCTCAACCTGACCG.........ACAACCGCAGCACCGGACAAC 550
       :::::::::|||   |||          ||||||::::::|||
 575 hrGlyThrThrThrValProValProThrThrThrGlySerPro...... 589

 551 GGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTA 600
               |||   :::|||   |||:::
 590 ......ThrThrGlnThrThrAlaProValThrLysProThrValProSe 604

 601 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATC 650
      :::|||   ::::::||||||||||   |||:::|||:::
 604 rSerThrThrGlnThrAlaProProValThrThrProThr.......... 617

 651 GGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACA 700
       ::::::||||||      :::|||:::|||      :::
 618 ..SerGlnProProValThrThrThrSerLeuLeuThrThr......... 630

 701 TCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGT 750
                           |||:::::::   :::::::|||
 631 ....................LeuThrThrProThrValProVal..... 638

 751 ATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTC...TGCT 797
               :::   |||:::::      :::|||   |||
 639 .........ThrThrThrValValProSerSerAlaThrValProThrT 652
```

```
    798 TTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGC 847
    652 hrProProThrThrValThrValAlaAlaThrThrThrSer..... 665

    848 AACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAAC 897
    666 LysAlaProValValThrThrSerProThrLeuAlaProThrSerProTh 682

    898 CCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGC 947
    682 rLysLeuProThrSer......ProProSerThrValGly.....T 694

    948 CATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCG 997
    694 hrSerProThrAlaProAlaAsnLeuThrThrProThrThrAla..... 708

    998 GCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTG 1047

    708 ..................................................... 708

   1048 CCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGC 1097
    709 ....ProValAsnProThrSerSerThrThrAlaProThrAlaProValA 724

   1098 CAAATACCCGTCCCCTTACCATTCCCGAAATA..... 1129
    724 snProThrSerProThrThrAlaProThrValProProValThrThrThr 740

   1130 .....TCCGTTCAAACTTGGAGCAGCGTTAC 1155
    741 ThrProThrThrThrThrSerThrThrThrThrThrThrThrThrTh 757

   1156 GGCAAAGAAAACATCACCTCCTCAACCG 1183
    757 rThrThrGlnThrThrProThrThrPro 766
```

```
seq_name: pir2:T47182

seq_documentation_block:
hypothetical protein DKFZp434M1616.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47182
R;Koehrer, K.; Beyer, A.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A;Reference number: Z24378
A;Accession: T47182
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1299 <AAA>
A;Cross-references: EMBL:AL162004
A;Experimental source: adult testis; clone DKFZp434M1616
C;Genetics:
A;Note: DKFZp434M1616.1


alignment_scores:
            Quality:  142.00              Length:    587
              Ratio:  0.568                Gaps:     29
 Percent Similarity:  42.589   Percent Identity: 21.465

alignment_block:
 US-09-303-518D-463 x T47182   ..

 Align seg 1/1  to: T47182  from: 1  to: 1299

      74 CAGATTTGG.............................CAAACGAT 90
     254 GlnValTrpAsnLysLysAsnAlaAsnGluLysGlyArgSerGlnThrSe 270

      91 CCCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAA 140
     270 rLysLeuProProArgPheAlaLys.........LysGlnAlaThrGlyI 284

     141 ATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCATA 190
     284 leGlnGlnAlaGlnSerSerAlaSerValProProLeuAlaSerAlaPro 300

     191 TCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGATTCAA 240
     301 LeuProProSerThrSerAlaSerValProAlaSerThr........... 313

     241 CAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCA 290
     314 .SerAlaProLeu.........ProAlaThrLeuThrProValProAlaS 327

     291 CGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATT...... 334
     327 erThrSerAlaProValProAlaSerThrLeuAlaProValLeuAlaSer 343

     335 .........CCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGC 375
     344 ThrSerAlaProValProAlaSerProLeuAlaProValSerAlaSerAl 360

     376 C...TTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGG 422
     360 aSerValSerAlaSerValProAlaSerThrSerAlaAlaAlaIleThrS 377

     423 CTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGG 472
     377 erSerSerAlaProAlaSerAlaProAlaProThrProIleLeuAlaSer 393

     473 ATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCC........ 514
     394 ValSerThrProAlaSerValThrIleLeuAlaSerAlaSerIleProIl 410

     515 ................GCCTCAACCTGACCGACAACCGCAGCACCGGACA 548
     410 eLeuAlaSerAlaLeuAlaSerThrSerAlaProThrProAlaProAlaA 427

     549 ACGGCTTGCCG..............ACCGTTTCCACAATGCCGGCGCTA 583
     427 laSerSerProAlaAlaProValIleThrAlaProThrIleProAlaSer 443

     584 TGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCA........... 622
     444 .................AlaProThrAlaSerValProLeuAlaProAl 454

     623 .............CCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGC 659
     454 aSerAlaSerAlaProAlaProAlaProThrProValSerAlaProAsnP 471

     660 CGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCG 709
     471 roAlaProProAlaProAla...GlnThrGlnAlaGlnThrHisLysPro 486

     710 CGGCAGGAGAAATTG.................TCGGCGCAGGCGATGCC 741
     487 AlaGlnAsnProLeuGlnThrThrSerGlnSerSerLysGlnProProPr 503

     742 GTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGG 791
     503 oSerIleArgLeuProSerAlaGlnThr..................... 512

     792 TCTGCTTTCCACCGAAAACAAGATGGCGCGCATCA............... 826
     513 ........ProAsnGlyThrAspTyrValAlaSerGlyLysSerIleGln 526

     827 ...................ACGATTTGGCAGATATGGCGCAACTCAAA 855
     527 ThrProGlnSerHisGlyThrLeuThrAlaGluLeuTrpAspAsn...Ly 542

     856 GACTATGCCGCAGCAGCCATCCGCGATTGGGCAG...TCCAAAACCCCAA 902
     542 sValAlaPro......ProAlaValLeuAsnAspIleSerLysLysLeuG 557
```

```
 903 TGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCC 952
     |||   :::      :::|||   :::::|||   |||:::     |
 557 lyProIleSerProProGlnProProSerValSerAlaTrpAsnLys..P 573

 953 CCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGC..... 990'
     ||:::   :::::::|||:::::::   ::::::   |||
 573 roLeuThrSerPheGlySerAlaProSerSerGluGlyAlaLysAsnGly 589

 991 .................TTGGGCGGCATCACGGCACATCCTGTCAAGCG 1022
                         :::|||      |||
 590 GlnGluSerGlyLeuGluIleGlyThrAspThrIle..... 601

1023 GTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCG 1072
         |||:::|||||||   |||        :::::::::   |||
 602 ....GlnPheGlyAlaProAlaSerAsnGlyAsnGluAsnGluValValP 617

1073 ACAATTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCC 1122
          :::::::::   :::      |||   |||:::|||
 617 roValLeuSerGluLysSerAlaAspLysIleProGluPro..... 630

1123 CGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCAC 1172
                           |||||||||
 631 .................LysGluGlnArg..... 634

1173 CTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACC 1222
                                                      |
 635 ..............................................G 635

1223 AACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT 1272
     ||:::::::|||:::::::|||   |||:::   ::::::   :::|||:::
 635 lnLysGlnProArgAlaGly...ProIleLysAlaGlnLysLeuProAsp 650

1273 TTT.....................................GA 1277
     :::                                    ||
 651 LeuSerProValGluAsnLysGluHisLysProGlyProIleGlyLysGl 667

1278 GAAGCACGTGAAATATGATACGAAGCTCGATATTCAAGAATTATCGGGGG 1327
     |:::   :::|||              |||   |||:::::::::
 667 uArgSerLeuLysAsnArgLysValLysAspAlaGlnGlnValGluProG 684

1328 GCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAG 1377
       |||      |||   :::|||:::                      .
 684 luGlyGlnGluLysProSerProAla..... 692

1378 GTTGATAGGAAGCTTAATAAATTGACAACTCGTGAGCAGGTGGAGAAAAA 1427
     :::   |||:::   :::   :::|||||||:::|||          |||
 693 ThrValArgSerThrAspProValThrThrLysGlu......ThrLysAl 707

1428 TGTTCAGGAA.................ACGAGAAGAAGGAGTCAGAGTA 1459
      |||:::|||                |||          :::|||:
 707 aValSerGluMetSerThrGluIleGlyThrMetIleSerValSerSerA 724

1460 GTCAGTTTAAAGCCCATGCGCAACGAGAA.............TGGGAA 1494
     ::::::::   :::::::|||:::   |||              ||||||
 724 laGluTyrGlyThrAsnAlaLysMetGluSerAlaArgLysAlaTrpGlu 740

1495 AATAAAACAGGGTTAGATTTTAATCATTTTATAGGTGGTGATATCAATAA 1544
     |||:::                              ::::::::::
 741 AsnSerPro.........................AsnValArgGl 747

1545 GAAAGGCACA 1554
     :||||||:::
 747 uLysGlySer 750
```

seq_name: pir2:T30826

seq_documentation_block:
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N;Alternate names: alpha-NAC protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 24-Nov-1999
C;Accession: T30826
R;Yotov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A;Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a mus
A;Reference number: Z20889; MUID:96312450
A;Accession: T30826
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2187 <YOT>
A;Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1
C;Genetics:
A;Gene: Naca
A;Map position: 10
A;Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A;Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding ac
C;Keywords: alternative splicing; DNA binding; transcription factor

alignment_scores:
Quality: 141.50 Length: 569
Ratio: 0.546 Gaps: 27
Percent Similarity: 45.518 Percent Identity: 22.847

alignment_block:
US-09-303-518D-463 x T30826 ..

Align seg 1/1 to: T30826 from: 1 to: 2187

```
 254 AAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTC 303
     |||:::   |||||||||||||||         :::   ::::::::::||
 833 LysAspThrSerAlaThrLeuSerLeuLysSerValProAlaValThrSe 849

 304 CATTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGCCGG 353
     |:::   |||      :::                              ||||
 849 rLeuSerProProLysAla.........................ProV 857

 354 TAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACG 403
     ||:::|||   :::::::::                        |||   :::
 857 alAlaProSerAsnGluAla.....................ThrIleVal 866

 404 AACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCC 453
       |||      ||||||::::::   :::      |||||||||   ||
 867 ProThrGluIleProThrSerLeuLysAsnAlaLeuAlaAlaAlaThrPr 883

 454 GCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGC 503
     |                                  |||         ::::::|
 883 o.......................LysGluThrLeuAlaThrSerIleP 892

 504 CCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGC 553
     |||||:::                   ||:::   :::|||::::::
 892 roLysVal.................ThrSerProSerProGlnLysThr 902

 554 TTGCCGACCGTTTCCACA......ATGCCGGCGCTATGCTGACGCAAGGA 597
           :::||||||          ||||||:::      :::|||
 903 ProLysSerValSerLeuLysGlyAlaProAlaMetThrSerLysLys.. 918

 598 GTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAG 647
       ||||||:::      |||      |||::::::||||||
 919 ....AlaThrGluIleAlaAlaSerLysAspValSerProSerGlnPheP 934

 648 ATCGGGCAATGCCG..............CCGAAGCCTTCAACGGCACTG 682
       :::   :::|||               |||      |||
 934 roLysGluValProLeuLeuGlnHisValProProThrSerProProLys 950

 683 CAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA 732
     :::      |||   |||   |||:::         :::         ::
 951 SerProValSerAspThrLeuSerGlyAlaLeuThrSerProProProLy 967

 733 GGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCA 782
     ::::   |||:::   :::   |||::::::   |||
 967 sGlyProProAlaThrLeu...AlaGluThrProThr..... 978

 783 CGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATT 832
                    :::|||   |||:::         |||:::
 979 .............TyrProLysLysSerProLysProAlaAla..... 988
```

```
 833 TGGCAGATATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGAT 882
                      ||||||   |||    ||||||
 989 ..................SerLysLysThrProAlaThrProSerProGl 999

 883 TGGGCAGTCCAAAACCCCAATG..............CCGCACAAGGCAT 917
      |||   :::   :::|||:::                 ||||||
 999 uGlyValThrAlaValProLeuGluIleProProCysSerLysLysAlaP 1016

 918 AGAAGCCGTCAGCA.............................. 931
       |||   :::|||
1016 roLysThrAlaAlaProLysGluSerSerAlaThrSerSerSerLysArg 1032

 932 .........ATATCTTTATGGCAGCCATCCCCATCAAAGGGA........ 964
             ::::::   :::      ||||||||||||
1033 AlaProLysThrAlaValSerLysGluIleProSerLysGlyValThrAl 1049

 965 .......TTGGAGCTGTCC............................. 976
           ||||||:::|||
1049 aValProLeuGluIleSerLeuProLeuLysGluThrSerLysSerAlaT 1066

 977 .....GGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGC 1021
       ||||||::::::|||                          ||||||
1066 hrProGlyGluLysSerAla.....................SerSer 1074

1022 GGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGC 1071
     :::|||      :::::         |||   :::||||||:::::
1075 ProLysArgSerProLysThrAlaGlyProLysGluThrProProGlyGl 1091

1072 GACAATT.....................TTGCCGATGCGGCATACGCCAA 1100
     ::::                       ||||||   ::::::||||||:
1091 yValThrAlaValProProGluIleSerLeuProProLysGluThrProG 1108

1101 ATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGC 1150
     :::::         :::   |||:::   |||:::   :::   |||:::
1108 lnAsn.........AlaThrProAsnGluSerLeuAlaAlaSerSerGln 1121

1151 GTTACGGCAAAGAAAACATCACCTCCTCAACC.GTGCCGCCGTCAAACGG 1199
          :::   |||||||||   |||:::   :::||||||::::::
1122 LysArgSerProLysThrSerValProLysGluThrProProGlyGlyVa 1138

1200 CAAAAATGTCAAACTG............GCAGACCAACGCCACCCGAAGA 1237
           :::   |||            |||   |||:::   |||||||
1138 lThrAlaMetProLeuGluIleProSerAlaProGlnLysAlaProLysT 1155

1238 CAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAATTTTGAGAAGCACGTG 1287
     ||:::||||||         |||   :::|||:::   |||      |||
1155 hrAlaValPro.........LysGlnIleProThrProGluAspAlaVal 1168

1288 AAATAT......GATACGAAGCTCGATATTCAAGAATTATCGGGGGGCGG 1331
                  :::   |||:::   ::::::   |||       ::
1169 ThrIleLeuAlaGlySerProLeuSerProLysLysAlaSerLysThrAl 1185

1332 TATACCTAAGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAGGTTG 1381
     :   ||||||      |||:::      :::                |||:
1185 aAlaProLysGluAlaProAlaThrProSerValGlyValIleAlaValS 1202

1382 ATAGGAAGCTTAAT.........AAATTGACAACTCGTGAGCAGGTGGAG 1422
     ::   :::::::::         |||   |||:::::::      :::
1202 erGlyGluIleSerProSerProLysLysThrSerLysThrAlaAlaPro 1218

1423 AAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGAGTAGTCAGTTTAAAGC 1472
     |||:::   ::::   |||         ::::::|||   :::      ||
1219 LysGluAsnSerAlaThrLeuProProLysArgSerProLysThrAlaAl 1235

1473 CCATGCGCAACGAGAATGGGAAAATAAAACAGGGTTAGATTTTAATCATT 1522
     |       :::             ::::::   |||:::
1235 aProLysGluThrProAlaThrSerSerGluGlyValThr.........A 1249

1523 TTATAGGTGGTGATATCAAT.....................AAGAAAGGC 1551
       :::   ::::::|||:::                    :::||||||
1249 laValProSerGluIleSerProSerProProThrProAlaSerLysGly 1265

1552 .........ACAGTAACA.........GGAGGGCATAGTCTAACCCGTGG 1583
                ||||:::|||         :::   ::::::|||:::::::
1266 ValProValThrLeuThrProLysGlyAlaProAsnAlaLeuAlaGluSe 1282

1584 TGATGTA......CGGGTGATACAACAAACCTCGGCACCTGATAAACATG 1627
     :   :::       :::   :::   :::|||:::||||||::::::   :
1282 rProAlaSerProLysLysValProLysThrAlaAlaProGluGluThrS 1299

1628 GGGTTTATCAAGCGACAGTGGAAATTAAAAAGCCTGATGGAAGTTGGGAG 1677
     :::::       :::       :::|||   |||       |||      |||
1299 erThrThrProSerProGlnLysIleProLysValAlaGlyProLysGlu 1315

1678 GTGAAAACGAAAAAAGGTGGGAAAGTGATGACCAAGCACACCATGTTCCC 1727
     :::::::::          :::|||      |||      |||      ||
1316 AlaSerAlaThrProProSerLys...LysThrProLysThrAlaValPr 1331

1728 AAAAGAT 1734
     |||:::
1331 oLysGlu 1333
```

seq_name: pir2:S35047

seq_documentation_block:
mucin JUL7 - human
C;Species: Homo sapiens (man)
C;Date: 10-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 08-Oct-1999
C;Accession: S35047
R;Dufosse, J.; Porchet, N.; Audie, J.P.; Guyonnet Duperat, V.; Laine, A.; van-Seuning
Biochem. J. 293, 329-337, 1993
A;Title: Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic altern
A;Reference number: S35047; MUID:93343858
A;Accession: S35047
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-543 <DUF>
A;Cross-references: EMBL:X74370; NID:g407081; PIDN:CAA52408.1; PID:g407082
A;Note: the authors translated the codon AAA for residue 63 as Gln and CGG for residu
A;Note: the amino acid sequence from Fig. 2 is inconsistent with the nucleotide seque

alignment_scores:
Quality: 141.00 Length: 467
Ratio: 0.698 Gaps: 27
Percent Similarity: 43.255 Percent Identity: 24.625

alignment_block:
US-09-303-518D-463 x S35047 ..

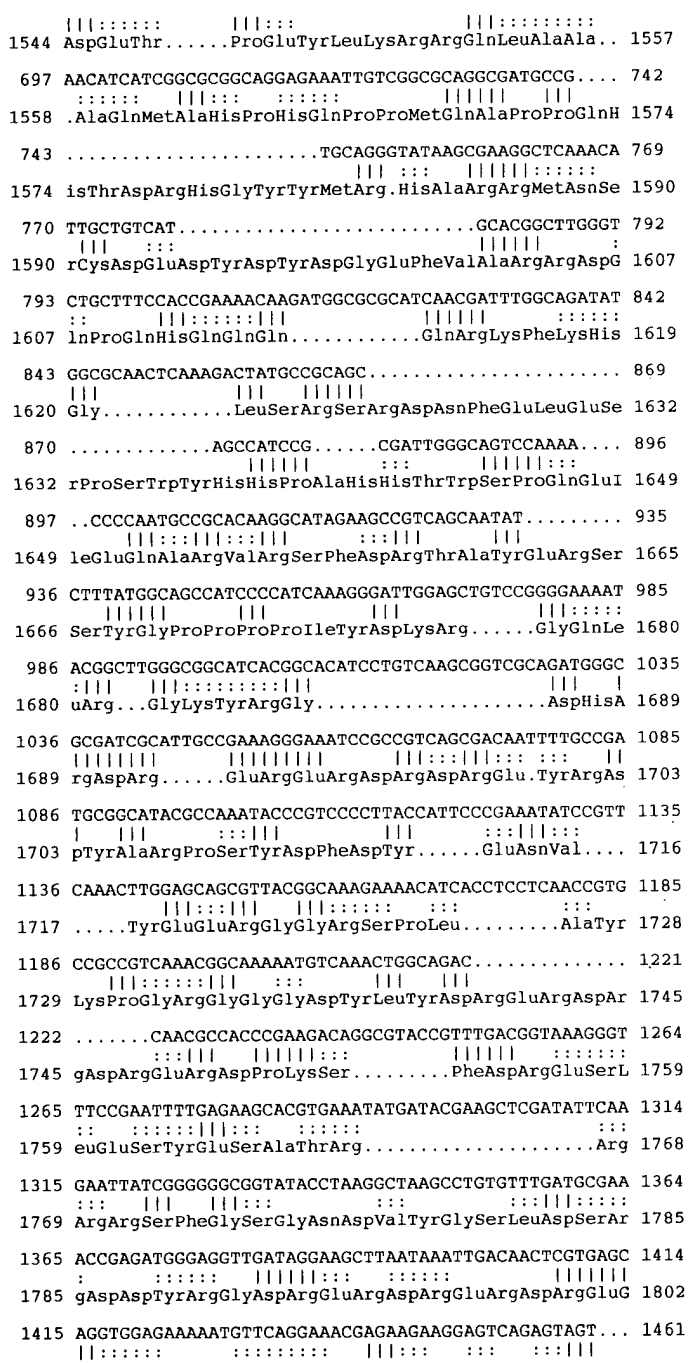
Align seg 1/1 to: S35047 from: 1 to: 543

```
  68 ACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCGT 117
     ||||||:::      :::|||                       |||||||
 138 ThrProGluThrThrHisThr.....................SerThrVa 147

 118 CAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGGGA 167
     |      ::::::   |||      ||||||   ::::::::::::   |
 147 lLeuThrThrLysAlaThr......ThrThrArgAlaThrSerSerThrS 162

 168 GCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAAGCCATC 217
     ||   ||||||:::|||      :::   |||      :::
 162 erThrProSerSerThrProGlyThrThrTrpIleLeuThrGluLeuThr 178

 218 AGTTGGGCCACCTGATGATTCAACAGGCGGCCG..................  250
     :::   ||||||         :::      |||
 179 ThrAlaAlaThrThrThrAlaGlyThrGlyProThrAlaThrProSerSe 195

 251 .....................TTGAAGGAAATATCGGCTACATTGTCCG 278
                          |||   |||::::::::::|||   ::::
 195 rThrProGlyThrThrTrpIleLeuThrGluLeuThrThrThrAlaThrT 212

 279 CTTTTCCGATCACGGGC.........ACAAATTCCATTCGCCCTTCGACA 319
     ::         ||||||         |||   |||      |||:::|||
 212 hrThrAlaSerThrGlySerThrAlaThrLeuSerSerThrProGlyThr 228
```

```
 320 ACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGA 369
229 ThrTrp............IleLeuThrGluProSerThrThrAlaThrVa 241

 370 TTCAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGA 419
241 lThrValProThrGlySerThrAlaThrAlaSerSerThr...GlnAlaT 257

 420 CGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGA 469
257 hrAlaGlyThrProHisValSerThr..................... 265

 470 GGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTC 519
266 .........ThrAlaThrThr...........ProThrValThrSerSe 275

 520 AACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTT.... 565
275 rLys...ProLeuLeuLeuSerProGlyLeuGlnLeuProSerAlaLeuA 291

 566 ..TCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCA 613
291 rgSerThr.............................AlaThrAsnThr 297

 614 AAC..............GCGCCACCCGATACAGCCCCG......AGCTG 642
298 HisSerTyrGlnLeuTyrSerHisProLeuLeuProGlyThrThrTr 314

 643 GACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGT 692
314 pThrArgLeuSerGlnThrThrThrProThrAlaThrMet......SerT 329

 693 CAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCG 742
329 hrAlaThrProSerSerThrProGluThrVal.................. 339

 743 TGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCT..... 787
340 ..............HisThrSerThrValLeuThrAlaThrAlaThrTh 351

 788 ..................TGGGTCTGCTTTCCACCGAAAACAAGATGGC 818
351 rThrGlyAlaProAlaLeuTrpProProProProPro.ProGlnGlu... 366

 819 GCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAG 868
367 .......................GlnLeuThrLeuProLysCysArgLe 375

 869 CAGCCATCCGCGATTGGGCAGT...CCAAAACCCCAATGCCGCACA.... 911
375 uProGlnProArgLeuHisSerHisProLeuLeuGlnProArgAspGlyT 392

 912 .........................................AGGCATAGA 920
392 hrHisAlaSerSerValAspGlnHisAsnHisHisThrHisAsnHisHis 408

 921 AGCCGTCAGCAATATCTTTAT.....................GGCAGCCA 949
409 ThrHisAsnGlnTrpLeuHisGlyAspProLeuLeuHisProGlyAspHi 425

 950 TCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGC 999
425 sProHisArgGlnSer.......................AlaAspHisH 434

1000 ATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCC 1049
434 isHis..HisAsnCysAlaThrGlySerMetAlaThrProSerSerSerT 450

1050 GAAAG......GGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCAT 1093
450 hrGlnThrSerGlyThrProProSerLeuThrThrThrAlaThrThrIle 466

1094 ACG.....................CCAAATACCCGTCCCCTTACCATTCC 1122
467 ThrAlaThrGlySerThrThrAsnProSerSerThrProGlyThrThrPr 483

1123 CGAAATATCCGTTCAAACTTGGAGCAGCGTTACG..............G 1157
483 oIleProProValLeuThrSerThrAlaThrThrProAlaAlaThrSerS 500

1158 CAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATG 1207
500 erLysAlaThrSerSerSerSerPro......ArgThrAlaThrThrLeu 514

1208 TCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACG 1255
515 ProValLeuThrSerThrAlaThrLysSerThrAlaThrSerPheThr 530

seq_name: pir2:C39135

seq_documentation_block:
hypothetical protein 3 (gyrB region) - Haloferax sp.
C;Species: Haloferax sp.
C;Date: 11-Feb-1993 #sequence_revision 11-Feb-1993 #text_change 18-Nov-1994
C;Accession: C39135
R;Holmes, M.L.; Dyall-Smith, M.L.
J. Bacteriol. 173, 642-648, 1991
A;Title: Mutations in DNA gyrase result in novobiocin resistance in halophilic archae
A;Reference number: A39135; MUID:91100352
A;Accession: C39135
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-437 <HOL>
A;Cross-references: GB:M38373


alignment_scores:
           Quality:  139.00            Length:    439
             Ratio:  0.785               Gaps:     25
 Percent Similarity:  40.319   Percent Identity: 25.285

alignment_block:
 US-09-303-518D-463 x C39135   ..

 Align seg 1/1  to: C39135  from: 1  to: 437

  102 GCAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTAT 151
   17 AlaArgAlaArgProArgThrArgArgArgHisArgAsnAspHisProLe 33

  152 TCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCATAT.......... 191
   33 u.....................LeuAlaGlyArgArgTyrLeuArgAspA 43

  192 ........CGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGAT 233
   43 spArgValArgLeuGlnAspAlaArgLysProProAla............ 55

  234 GATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTT 283
   56 ..............................ArgAlaArgValProGlnLe 62

  284 CCGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACAT 333
   62 uArgGlyArg.........AspPheAlaLeuArgArg.............. 71

  334 TCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCG 383
   72 .......................AlaAspArgArgValGluHisValPro 80

  384 ............CATCCATTGGGACGGATACGA................... 404
   81 LeuArgGlyArgHisPro......ArgValArgArgValProGlnArgAs 95

  405 ..........ACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGC 444
```

```
                ::::|||   ||||||:::|||   |||          ||||
 95 pGlnAspGlyAlaProArgArgArgHisLeuLeuArg.........ArgA 109

445 GGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACAT... 491
    ||:::::::      :::   |||:::               ||||||
109 rgValGlyGlyHisArgGlyArgAsn..............ArgHisAla 120

492 ...AAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCA 538
       ||||||:::|||            :::      :::|||   ||
121 GlyAspArgArgAlaProGlyValAspSerArgLeuArgGlnGlnHisGl 137

539 GCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTG 588
    ||||                  |||:::   |||:::   :::|
137 nHis.......................ProArgGlyArgHisAlaSerA 146

589 AC.......................GCAAGGAGTAGGCGACGGATTCAA 614
    ||                        |||:::|||   |||   |||
146 spArgValGlnAspGlyAlaHisProArgArgGlnArgLeuArgGluGln 162

615 ACGCGCCACCCGATACAGCCCCGAGCTGGACAG.........ATCGGGCA 655
       ||||||      :::|||||||      |||            |||::
163 ProArgHisAlaGlyArgProArgArgArgGlnProProArgArgGlyAr 179

656 ATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCATC 705
    :   |||   :::   :::||||||   |||          :::
179 gSerArgGlyThrHisArgArgHisLeuArgGlnAlaProArgProAlaV 196

706 GGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAG 755
      ||||||      :::   :::   |||:::   |||:::
196 alArgGlyProAspGluAspGlnAlaArgGluPheArgGlyProArgHis 212

756 CGAAGGCTCAAACATTGCTGTCATGCACGG.........CTTGGGTCTGC 796
    ||||||   :::|||         ||||||            |||:::
213 ArgArgGluArgHisProProThrAlaArgAspValLeuArgGlyGluPr 229

797 TTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCG 846
       |||         ||||||   |||
229 oGlyHis......GlyAspGlyHisHis..................... 236

847 CAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAA 896
                        :::   |||   |||   |||:::
237 .....................LeuGluGlyArgArgGlyArgProArg 245

897 CCCCAATGCCGCACAAGGCATAGAAGC.................CGTCA 928
    ||||||   |||         |||:::                 ||| :
246 ProGlnGlyArgGluAlaGlyArgGlyAlaHisProProGlnValArg.A 262

929 GCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATT............ 966
    :::::|||:::::::||||||         :::|||:::
262 laArgIleTyrLeuAlaAlaGlyGluAlaArgGlyLeuProGluProArg 278

967 ............................GGAGCTGTCCGGGGAAAATA 986
                                ||||   :::||||||:::
279 ProLeuGlyValArgThrValHisArgGlyGlyArgLeuArgGlyArgVa 295

987 CGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCG 1036
     |||   :::|||                  |||   |||:::   :
295 lGlyGlnAlaGly.................ProArgProGlnValProG 306

1037 CGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGAT 1086
    ::         |||:::|||:::                      |||
306 lyAspPheAlaProGlnGlyGlu.....................Asp 314

1087 GCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTC 1136
    :::          :::   |||         ||||||||||:::   |||::
315 SerGluArgArgGluThrProProArgProHisSerArgLysArgArgAs 331

1137 AAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACCGTGC 1186
    ::::  ||||||                  ||||||         |||
331 pThr..GlyAla.................HisHisArgHisTrpArgAr 341

1187 CGCCGTCAAACGGCAAAAATGTCAAACTGGCAGA......CCAACGCCAC 1230
           :::|||   :::   :::   |||           :::  |
```

```
 341 gArgArgArgArgValArgHisArgGluGlyAlaLeuProAlaAlaHisP 358

1231 CCGAAGACAGGCGT 1244
     ||:::|||||||||
 358 roAspAspArgArg 362

seq_name: pir2:A46299

seq_documentation_block:
tyrosine kinase substrate disabled - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 22-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 01-Aug-1997
C;Accession: A46299
R;Gertler, F.B.; Hill, K.K.; Clark, M.J.; Hoffmann, F.M.
Genes Dev. 7, 441-453, 1993
A;Title: Dosage-sensitive modifiers of Drosophila abl tyrosine kinase function: prosp
A;Reference number: A46299; MUID:93194063
A;Accession: A46299
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-2411 <GER>
A;Cross-references: GB:L08845
A;Note: sequence extracted from NCBI backbone (NCBIP:127163)
C;Keywords: alternative splicing; phosphoprotein


alignment_scores:
            Quality:  138.50              Length:    558
              Ratio:  0.523                 Gaps:     38
 Percent Similarity:  47.491   Percent Identity: 23.835

alignment_block:
 US-09-303-518D-463 x A46299   ..

 Align seg 1/1  to: A46299  from: 1  to: 2411

  266 GCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGC...CC 312
      ||||||  ||||||  |||      |||:::  :::::::::  ||
 1405 AlaThrSerSerAlaAlaProSerArgGlyAlaProGlyLeuHisThrPr 1421

  313 TTCG.............ACAACCATGCCTCACATTCCGATTCTGACGA 347
      ||||                ::::|||  |||  ||||||  |||  |
 1421 oSerGlnPheAsnAspValSerThrSerPro...IlePro...LeuGlnL 1436

  348 AGCCG.............GTAGTCCCGTTGACGGATTCAGCCTTTACC 382
      |||||                  ||||||        ||||||  :::
 1436 ysProGlyMetGlyProSerProValProSerGlnLeuSerAlaValSer 1452

  383 GCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGG 432
          |||  |||  |||:::                  |||  ||
 1453 GlnLeuIleAspThrAlaThrLys.............GlnMetMetGl 1464

  433 CCACAGGGCG.................GCGGCTATCCCGCTCCCAAAGG 464
      |  :::                    |||::::::      |||||||
 1464 yAspLysAspArgGluLysGlnSerTrpAlaThrPheAspSerProLysA 1481

  465 CGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCC 514
      ||:::|||          |||                |||      |||
 1481 laLysGlyLysAlaArgLeuThrLeuProProProProProAlaSer 1497

  515 GCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCG..... 559
         ::::::  |||  |||        :::|||  |||
 1498 AsnThrSerGlnProAspThr.........GluAsnArgLeuAlaValAr 1511

  560 .............ACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGG 596
                  |||::::::        ::::::      |||::::
 1511 gIleProGlyMetThrAlaGlyGlnSerAsnSerValValGlyArgArgA 1528

  597 AGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACA 646
      ::   ||||||  |||::::::      |||  :::|||  |||
 1528 rgSerAlaThrThrSerSerSerSerArgAspLeuSerPro...TrpAsp 1543

  647 GATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAA 696
```



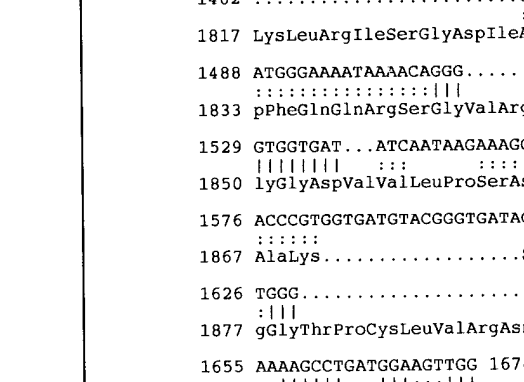
seq_name: pir2:B41206

seq_documentation_block:
microtubule-associated protein 4 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 05-Nov-1999
C;Accession: B41206
R;West, R.R.; Tenbarge, K.M.; Olmsted, J.B.
J. Biol. Chem. 266, 21886-21896, 1991
A;Title: A model for microtubule-associated protein 4 structure. Domains defined by c
A;Reference number: A41206; MUID:92042100
A;Accession: B41206
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1125 <WES>
A;Cross-references: GB:M72414; NID:g456554; PIDN:AAA16372.1; PID:g199025
C;Superfamily: MAP2/tau repeat homology
C;Keywords: microtubule binding; tandem repeat
F;904-934/Domain: MAP2/tau repeat homology <MT1>
F;960-972/Domain: MAP2/tau repeat homology #status atypical <MT2>
F;973-1003/Domain: MAP2/tau repeat homology <MT3>
F;1004-1034/Domain: MAP2/tau repeat homology <MT4>
F;1035-1072/Domain: MAP2/tau repeat homology #status atypical <MT5>

alignment_scores:
Quality: 138.00 Length: 632
Ratio: 0.502 Gaps: 22
Percent Similarity: 43.513 Percent Identity: 20.411

alignment_block:
US-09-303-518D-463 x B41206 ..

Align seg 1/1 to: B41206 from: 1 to: 1125



seq_name: pir2:T15316

seq_documentation_block:
hypothetical protein B0302.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

```
C;Accession: T15316
R;Du, Z.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid B0302.
A;Reference number: Z18328
A;Accession: T15316
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1223 <DUZ>
A;Cross-references: EMBL:U41032; NID:g1086728; PID:g1086729; PIDN:AAA82367.1; CESP:B0302
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:B0302.1
A;Introns: 50/3; 106/3; 153/3; 180/3; 243/3; 399/2; 438/3; 482/1; 520/2; 545/1; 589/1; 6

alignment_scores:
            Quality:  137.50            Length:    523
              Ratio:  0.593              Gaps:     24
 Percent Similarity:  44.359   Percent Identity: 22.753

alignment_block:
 US-09-303-518D-463 x T15316  ..

 Align seg 1/1  to: T15316  from: 1  to: 1223

      12 CCGCAAAATATCCCTTATTCTGTCCATACTGGCAGTGTGCCTGCCGATGC 61
    715 ProAlaAsnIleProCysLeuValProThrProAlaProProAlaProAl 731

      62 ATGCACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTC 111
    731 a..HisPheSerGln.............ProValSerSerGlnArgVa 742

     112 GACCGTCAGCATTTCGAAC.............................. 130
    742 lAlaGlnGlnGlnGlnAsnThrLeuGlnLysAlaLeuAsnAspGluLeuL 759

     131 ...................CCGACGGGAAATACCACCTATTCGGCAGCA 160
    759 ysGlyAsnLeuAsnLysArgProThrGlyThrThrAlaProProSerAsn 775

     161 GGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAA 210
    776 GlyPheAsnAlaProArgAlaAspValAlaPro................. 786

     211 AGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAA 260
    787 .........................ValGlnGlnArgProIleSerSerA 795

     261 TATCG...GCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCC... 304
    795 laSerIleProAlaLeuGlnProGlnProIleGlnHisIleGlnLysPro 811

     305 ATTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGCCGGT 354
    812 IleGlnProGlnGlnVal......ArgIleProProSerThrAlaProVa 826

     355 AGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGA 404
    826 l............GlnLysProValGlnValSerAlaProThrHisSerA 839

     405 ACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCG 454
    839 snValAlaProThrThrSerSerGlnAlaSerAlaAspAlaArgAsnPro 855

     455 CTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCC 504
    856 LeuProPro.........................LysThrSerPr 862

     505 CAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCT 554
    862 oProValSerAsnThr......ProIleThrValAlaProValHisAlaA 877

     555 TGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCG 604
    877 laProThrThrSer.................................. 881

     605 ACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATCGGGC 654
    882 ..........................AlaProSerThrSerValValTh 889

     655 AATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACATCAT 704
    889 rArgArgProThrSerThrThrAla...GlnMetSerAspGluGluArgA 905

     705 CGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAA 754
    905 rgSerArgIleAlaMetAspIleSerSerAlaLeuPro............ 917

     755 GCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACC 804
    918 AlaProSerAlaLeuLeuTyrGlySerAsnSerThrSerSerLeuProSe 934

     805 GAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCA. 853
    934 rAlaAlaValSerThrAlaSerSerValProSerThrAlaArgAspAsnP 951

     854 .....AAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACC 898
    951 roValGluThrArgProSerGlnProHisVal..............Thr 962

     899 CCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCT...TTATGGCA 945
    963 MetProProLysLysSerSerGluProIleLeuSerSerGluValLeuGl 979

     946 GCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGG 995
    979 nProThrArgLeuProSerAlaThrThrSerGlnAlaLysProValThrG 996

     996 CGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCAT 1045
    996 lnProIleArgHis.................................. 1000

    1046 TGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGAT.GCGGCATA 1094
    1001 .........ProSerProProValAlaThrValIleProThrAlaValVa 1014

    1095 CGCCAAATACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGG 1144
    1014 lAspLysLysProVal.........SerGlnAsnGlnGlySerAsnValP 1028

    1145 AGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAACC......GTGCCG 1188
    1028 ro............LeuPheAsnIleThrAsnSerSerAsnGlyTyrPro 1040

    1189 CCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGAC 1238
    1041 GlnLeuAsnGly.........................TyrProAsnTy 1048

    1239 AGGCGTACCGTTTGACGGT.......................AAAGGGT 1264
    1048 rGlyAsnGlyPheGlnAlaTyrGlyTyrGlyMetAsnTyrHisGlnGlyT 1065

    1265 TTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCAA 1314
    1065 yrProGlyTyrGlnGlyTyrAsnSerTyrGlyAsnGlyMetGlyGlnLeu 1081

    1315 GAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAA 1364
    1082 AlaLeuThrHisAsnAlaValThrSerLeuProProLeuValProSerGl 1098

    1365 ACCGAGATGGGAGGTTGATAGGAAGCTTAATAAATTGACAACTCGTGAGC 1414
    1098 uAsnArgPheSerGly...................ThrAlaGlnProL 1108
```

```
  1415 AGGTGGAGAAAAATGTTCAGGAA.........ACGAGAAGAAGGAGTCAG 1455
            |||::::::::::::|||         |||::::::|||:::
  1108 euGlyGluSerAspIleMetGluPheLeuGlyThrGlnGlnArgGlnAla 1124

  1456 AGTAGTCAGTTTAAAGCC 1473
       :::|||:::   :::|||
  1125 GlySerSerSerArgAla 1130
```

seq_name: pir2:S51342

seq_documentation_block:
verprolin - yeast (Saccharomyces cerevisiae)
N;Alternate names: prolin-rich protein VRP1; protein L8300.13; protein YLR337c
C;Species: Saccharomyces cerevisiae
C;Date: 23-Feb-1995 #sequence_revision 11-Aug-1995 #text_change 23-Mar-2001
C;Accession: S51342; S39626; S57435
R;Du, Z.
submitted to the EMBL Data Library, December 1994
A;Description: The sequence of S. cerevisiae cosmid 8300.
A;Reference number: S51339
A;Accession: S51342
A;Molecule type: DNA
A;Residues: 1-817 <DUZ>
A;Cross-references: EMBL:U19028; NID:g609380; PID:g609392; MIPS:YLR337c
R;Donnelly, S.F.H.; Pocklington, M.J.; Pallotta, D.; Orr, E.
Mol. Microbiol. 10, 585-596, 1993
A;Title: A proline-rich protein, verprolin, involved in cytoskeletal organization and ce
A;Reference number: S39626; MUID:95058201
A;Accession: S39626
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-307,'R',309-349,'R',351-688,'E',690-709,'HLRWIPVPLIAPVKTLNNGYFLQVDRRCNTSII
A;Cross-references: EMBL:Z26645; NID:g414785; PIDN:CAA81388.1; PID:g439289
R;Munn, A.L.; Stevenson, B.J.; Geli, M.I.; Riezman, H.
submitted to the EMBL Data Library, June 1995
A;Description: end5, end6, and end7: mutations that cause actin delocalization and block
A;Reference number: S57435
A;Accession: S57435
A;Molecule type: DNA
A;Residues: 1-162,'F',164-817 <MUN>
A;Cross-references: EMBL:X87806; NID:g871534; PID:g871535
A;Experimental source: strain W303
C;Genetics:
A;Gene: SGD:VRP1; MDP5; END5
A;Cross-references: SGD:S0004329; MIPS:YLR337c
A;Map position: 12R
F;5-14/Region: proline-rich
F;77-85/Region: proline-rich
F;114-180/Region: proline-rich
F;216-245/Region: proline-rich
F;305-336/Region: proline-rich
F;349-357/Region: proline-rich
F;372-382/Region: proline-rich
F;396-406/Region: proline-rich
F;421-445/Region: proline-rich
F;518-528/Region: proline-rich
F;567-577/Region: proline-rich
F;608-621/Region: proline-rich
F;649-661/Region: proline-rich
F;678-685/Region: proline-rich
F;704-710/Region: proline-rich

alignment_scores:
Quality: 137.00 Length: 372
Ratio: 0.878 Gaps: 18
Percent Similarity: 41.935 Percent Identity: 23.656

alignment_block:
US-09-303-518D-463 x S51342 ..

Align seg 1/1 to: S51342 from: 1 to: 817

```
  275 TCCGCTTTTCCGATCACGGGCACAAATTCCATTCGC...CCTTCGACAAC 321
      ||||||         :::||||||   |||   :::   ||||||   ::
   64 SerAlaSerGlySerSerGlyThrValSerSerLysGlyProSerMetSe 80

  322 CATGCCTCACATTCCGATTCTGACGAAGCCGGTA..... 355
      :   |||   ||||||   :::      |||
   80 rAlaProProIleProGlyMetGlyAlaProGlnLeuGlyAspIleLeuA 97

  356 ........GTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACG 397
              :::|||                          |||::::::
   97 laGlyGlyIleProLys.............LeuLysHisIleAsnAsn 108

  398 GATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGC 447
      ::::::::::|||   |||      :::               |||:::||
  109 AsnAlaSerThrLysProSerProSer.............AlaSerAl 120

  448 TATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGG 497
      |   |||:::|||                                   |
  120 aProProIlePro.............................GlyA 126

  498 CGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGAC 547
      ||:::|||:::::::::|||         |||          ||||||
  126 laValProSerValAlaAlaPro......ProIleProAsnAlaProLeu 140

  548 AACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGA 597
      :::      |||:::|||   ::::::|||:::
  141 SerProAlaProAlaValProSerIleProSer..... 151

  598 GTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAG 647
                  |||:::|||||||||         |||:::      :
  152 .............SerSerAlaProPro......IleProAspIleProS 162

  648 ATCGGGCAATGCCGCCG.................AAGCCTTCAACGGCA 679
      ::   |||   ||||||                 :::|||:::
  162 erSerAlaAlaProProIleProIleValProSerSerProAlaProPro 178

  680 CTGCAGATATCGTCAAAAACATCA..... 703
      |||   :::|||:::   ::::::
  179 LeuProLeuSerGlyAlaSerAlaProLysValProGlnAsnArgProHi 195

  704 .........................TCGGCGCGGCAGGAGAAATTGTCGG 728
                                  ::::::|||:::|||   |||
  195 sMetProSerValArgProAlaHisArgSerHisGlnArgLysSerSerA 212

  729 CGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCA 778
          ::::::|||
  212 snIleSerLeuProSer.............................. 217

  779 TGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAAC 828
                  |||      ||||||
  218 ............ValSerAlaProProLeuPro.............. 224

  829 GATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCG 878
                           |||   ::::::|||   :::   |||
  225 ......................SerAlaSerLeuProThrHisValSerA 234

  879 CGATTGGGCAGTCCAAAACCCCAATGCCG..................... 907
             |||:::      |||   |||
  234 snProProGlnAlaProProProProProThrProThrIleGlyLeuAsp 250

  908 CACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCC..... 952
         |||      |||||||:::      ::::::      ||||||
  251 SerLysAsnIleLysProThrAspAsnAlaValSerProProSerSerGl 267

  953 ....CCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGG 998
          |||:::   ||||||   :::   :::|||   :::|||      :
  267 uValProAlaGlyGlyLeuProPheLeuAlaGluIleAsnAlaArgArgS 284

  999 CATCACGGCACATC.........CTGTCAAGCGGTCGCAGATGGGCGCGA 1039
      :::::|||                ::::||||||   :::      ::::::
  284 erGluArgGlyAlaValGluGlyValSerSerThrLysIleGlnThrGlu 300

 1040 TCGCATTGCCGAAAGGGAAATCCGCCG.........TCAGCGACAATTTT 1080
```

```
      :::|||         ::::::||||||         ||||||      ::
   301 AsnHisLysSerProSerGlnProProLeuProSerSerAlaProProIl 317

  1081 GCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATAT 1130
       :|||      |||:::|||         |||   :::   |||      |
   317 eProThrSerHisAlaProProLeuProProThrAlaProProProProS 334

  1131 CCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCA......... 1171
       ||:::   :::   :::|||:::      |||   ||||||
   334 erLeuProAsnValThrSerAlaProLysLysAlaThrSerAlaProAla 350

  1172 ...CCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGCA 1218
          ||||||   |||                |||   ||||||:::   ::
   351 ProProProProProLeu.........ProAlaAlaMetSerSerAlaSe 364

  1219 GACCAACGCCACCCGA 1234
       :||||||:::::::::
   364 rThrAsnSerValLys 369

seq_name: pir1:VGBEX1

seq_documentation_block:
glycoprotein X precursor - equine herpesvirus 1 (strain Ab4p)
C;Species: equine herpesvirus 1
A;Note: host Equus caballus (domestic horse)
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999
C;Accession: H36802
R;Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
submitted to GenBank, March 1992
A;Description: The DNA sequence of equine herpesvirus-1.
A;Reference number: A36805
A;Accession: H36802
A;Molecule type: DNA
A;Residues: 1-797 <TEL>
A;Cross-references: GB:M86664; NID:g330791; PIDN:AAB02506.1; PID:g330862
R;Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
Virology 189, 304-316, 1992
A;Title: The DNA sequence of equine herpesvirus-1.
A;Reference number: A41831; MUID:92295566
A;Contents: annotation; possible protein-coding frames
A;Note: neither amino acid nor nucleotide sequence is given
C;Genetics:
A;Gene: 71
C;Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein homo
C;Keywords: glycoprotein; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-797/Product: glycoprotein X #status predicted <MAT>
F;23-465/Region: serine/threonine-rich
F;489-797/Domain: equine herpesvirus 1 glycoprotein homology <EHG>
F;766-790/Domain: transmembrane #status predicted <TMN>
F;590/Binding site: carbohydrate (Asn) (covalent) #status predicted


alignment_scores:
            Quality:  136.50              Length:    569
              Ratio:  0.559                Gaps:     20
 Percent Similarity:  42.882   Percent Identity: 21.968

alignment_block:
 US-09-303-518D-463 x VGBEX1   ..

 Align seg 1/1  to: VGBEX1  from: 1  to: 797

     110 TCGACCGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGC 159
         ||||||:::::::   ::::::   |||:::::::::::::   |||::::::
      22 SerThrThrThrThrGluThrThrThrSerSerSerSerSerThrSerGlySe 38

     160 AGGGGGGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACA 209
         :|||   |||      |||:::|||                    |||   :
      38 rGlyGlnSerThrSerSerGlyThr.................ThrAsnS 49

     210 AAGCCATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAA 259
         :::::   |||   :::|||
```

```
      49 erSerSerSerProThrThr.............................. 55

     260 ATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCG 309
               |||   |||:::   |||   |||:::|||::::::
      56 SerProProThrThrSerSerSerProProThrSerThrHisThrSerSe 72

     310 CCCTTCG......ACAACCATGCCTCACATTCCGATTCTGACGAAGCCGG 353
          ||||||      :::|||:::                  |||:::
      72 rProSerSerThrSerThrGlnSerSerSerThrAlaAlaThrSerSerS 89

     354 TAGTCCCGTTGACGGATTCAGCCTTTACCGCA.........TCCATTGGG 394
         :::|||   |||   |||:::   |||:::          |||   :::
      89 erAlaProSerThrAlaSerSerThrThrSerIleProThrSerThrSer 105

     395 ACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGC 444
         |||:::|||:::|||      |||||||||   |||         :::::
     106 ThrGluThrThrThrThrThrProThrAlaSerThrThrThrProThrTh 122

     445 GGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAA 494
         ::::         |||   :::   :::       ||||||:::|||
     122 rThrThrAlaAlaProThrThrAlaAlaThrThrThrAlaValThrThrA 139

     495 AGGCG..................................................... 499
         |||
     139 laAlaSerThrSerAlaGluThrThrThrAlaThrAlaThrAlaThrSer 155

     500 ............................TTGCCCAAAATATCCGCCTC 519
                                                 ::::::::
     156 ThrProThrThrThrThrProThrSerThrThrThrThrThrAlaThrTh 172

     520 AACCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCA 569
         :|||   |||||||||||:::       :::          |||:::::|
     172 rThrValProThrThrAlaSerThrThrThrAspThrThrThrAlaAlaT 189

     570 CAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCG 619
         ||      |||                      |||:::   ::::::|||
     189 hrThrThrAla.........AlaThrThrThrAlaAlaThrThrThrAla 202

     620 CCACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCC 669
               ||||||   :::   |||      :::
     203 AlaThrThrThrAlaAlaThrThrThrThrAlaAlaThrThrThrAlaTh 219

     670 TTCAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGA 719
         :::||||||         ||||||   |||::::::|||
     219 rThrThrAlaAlaThrThrSerSerAlaThrThrAlaAlaThrThrThrA 236

     720 AATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACA 769
             ::::::   |||                  |||   :::   |||
     236 laAlaThrThrThrAlaAlaThrThrThr......AlaAlaThrThrThr 250

     770 TTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCG 819
               :::   ||||||   :::              |||
     251 AlaAlaThrThrThrAlaAlaThrThrThrGlySerProThrSerGlySe 267

     820 CGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAG. 868
          :::||||||                  |||   :::   |||:::
     267 rThrSerThrThrGlyAla............SerThrSerThrProSerA 280

     869 ................................CAGCCATCCGCGATT 883
                                        :::   ||||||
     280 laSerThrAlaThrSerAlaThrProThrSerThrSerThrSerAlaAla 296

     884 GGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAAT 933
         :::   :::::||||||   |||   :::|||          ||||||
     297 AlaThrThrSerThrProThrProThrSerAla...AlaThrSerAlaGl 312

     934 ATCTTTATGGCAG.....................................C 947
          |||      :::                                    |
     312 uSerThrThrGluAlaProThrSerThrProThrThrAspThrThrThrP 329

     948 CATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCG 997
         |||||   :::                |||   |||:::|||:::   :::
     329 roSerGluAlaThrThrAlaThrThrSerProGluSerThrThrValSer 345
```

```
  998 GCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTG 1047
      ||||||            ::::::            ::::::|||||
  346 AlaSerThrThrSerAlaThrThrThrAlaPheThrThrGluSerHisTh 362

 1048 CCGAAAGGGAAAT.....................CCGCCGTCAGCG. 1072
               :::                      ||||||:::
  362 rSerProAspSerSerThrGlySerThrSerThrAlaGluProSerSerT 379

 1073 .....ACAATTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACC 1117
           |||:::   |||         ||||||:::|||      :::
  379 hrPheThrLeuThrProSerThrAlaThrProSerThrAspGlnPhe... 394

 1118 ATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAAC 1167
                         |||   ||||||:::::::::::::::
  395 ....................ThrGlySerSerAlaSerThrGluSerAs 404

 1168 ATCACC....TCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAAC 1213
       |||      ||||||||||||||   ::::::   :::::::::
  404 pSerThrAspSerSerThrValProThrThrGlyThrGluSerIle.... 419

 1214 TGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGGG 1263
                 :::      :::|||                :::   |||
  420 ........ThrGluSerSerSerThrThrGluAlaSerThrAsnLeuGly 433

 1264 TTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATATTCA 1313
            ::::::|||:::            :::|||      |||   ::
  434 SerSerThrTyrGluSerThrGluAlaLeuGluThrPro...AspGlyAs 449

 1314 AGAATTATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGA 1363
      :      ||||||:::      |||:::   :::|||
  449 nThrThrSerGlyAsnThrThrProSerProSerPro.............. 461

 1364 AACCGAGATGGGAGGTTGATAGGAAGCTTAATAAATTGACAACTCGTGAG 1413
                           :::      :::::::::      :::
  462 ......................ArgThrProSerPheAlaAspThrGln 470

 1414 CAGGTGGAGAAAAATGTTCAGGAAACGAGAAGAAGGAGTCAGAGTAGTCA 1463
      |||:::      |||      :::|||::::::   :::   ::::::::
  471 GlnThrProAspAsnGlyValSerThrGlnHisThrThrIleAsnAspHi 487

 1464 GTTTAAAGCCCATGCGCAACGAGAATGGGAAAATAAAACAGGGTTAGATT 1513
      :      |||:::||||||::::::
  487 sThrThrAlaAsnAlaGlnLysHis.......................A 496

 1514 TTAATCATTTTATAGGTGGTGATATCAATAAGAAAGGCACAGTAACAGGA 1563
        :::|||      |||         :::::::::|||:::      |||
  496 laGlyHisHisArgGlyArgAlaGlyGlyArgArgGlySerProGlnGly 512

 1564 GGGCATAGTCTAACCCGTGGTGATGTACGGGTGATACAACAAACCTCGGC 1613
      |||         |||            |||:::         |||   ::
  513 GlySerHisThrThrProHisProAspArgLeu.........ThrProSe 526

 1614 ACCTGAT 1620
      :||||||
  526 rProAsp 528
```

seq_name: pir2:T51023

seq_documentation_block:
hypothetical protein B7F21.40 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C;Accession: T51023
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A;Reference number: Z25286
A;Accession: T51023
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2649 <SCH>
A;Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.40
A;Experimental source: BAC clone B7F21; strain OR74A

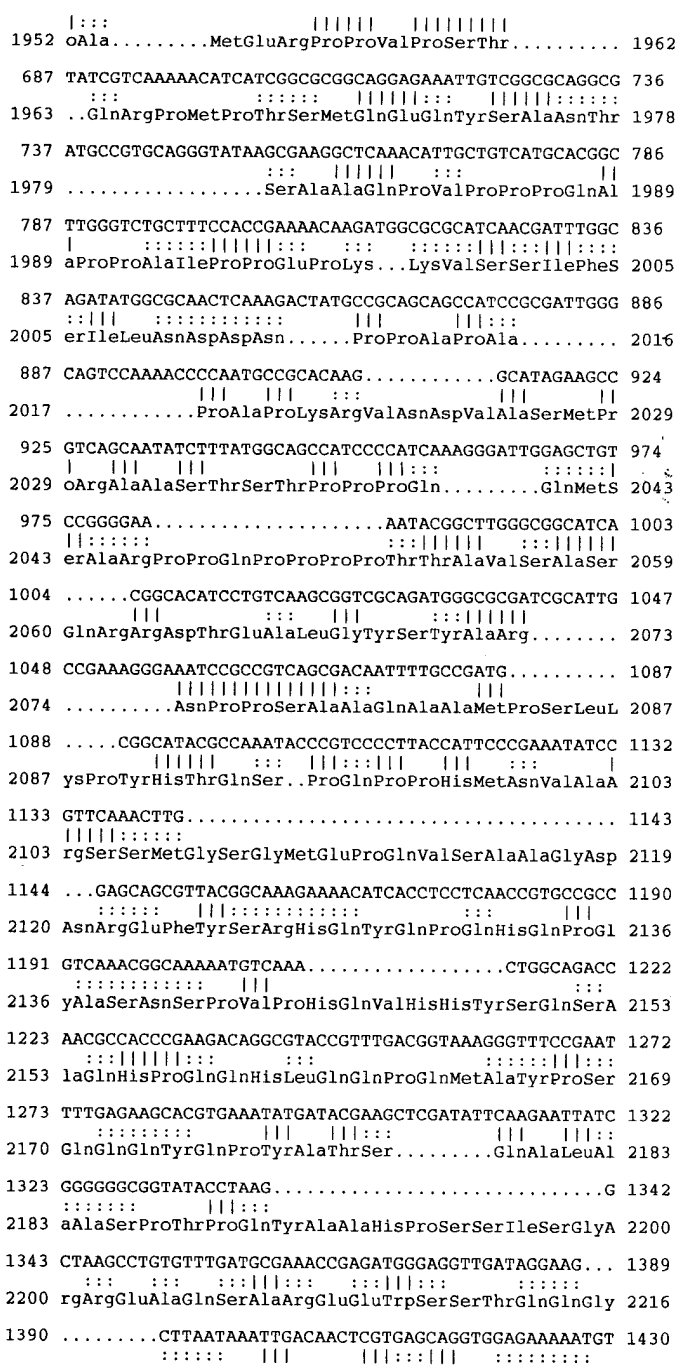
C;Genetics:
A;Gene: NCSP:B7F21.40
A;Map position: 6
A;Introns: 1619/3; 2584/1

alignment_scores:
Quality: 136.00 Length: 693
Ratio: 0.450 Gaps: 35
Percent Similarity: 43.579 Percent Identity: 19.769

alignment_block:
US-09-303-518D-463 x T51023 ..

Align seg 1/1 to: T51023 from: 1 to: 2649

```
   99 CCGGCAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAATACCA.. 146
      |||:::         |||::::::   :::|||         :::|||
 1747 ProSerArgProGlyProAlaValSerGlnThrMetSerProValProHi 1763

  147 .CCTATTCGGCAG..................................... 158
       |||:::||||||
 1763 sProLeuArgGlnProThrAlaThrPheValPheThrGluArgGluGlyG 1780

  159 .................CAGGGGGGAGCTTGCCNAGCGCAACGGCCATAT 191
                       |||      |||   :::   ::::::|||
 1780 luProIleProValSerGlnProProAlaGlnThrGlnHisGlnProVal 1796

  192 CGGATTGGGAAACATACAAAG......CCATCAGTTGGGCCACCTGATGA 235
      ||||||::::::   |||         ||||||   :::   |||:::
 1797 ArgIleSerGlnLysThrAlaProValProSerSerSerMetProSerAl 1813

  236 TTCAACAGGC..................................... 245
       |||   :::
 1813 aSerGluAlaMetProArgSerAlaGlyTrpIleProGluAlaAsnArgP 1830

  245 ..................................................... 245

 1830 roThrProLeuLeuSerGlnGlnHisGluLeuArgGluValArgGluIle 1846

  246 ......GGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATC 289
            ||||||   |||:::   |||::::::   ||||||   :::
 1847 ArgGluGlyArgGluArgGlnIleArgMetGlu...ProLeu...GlnPr 1861

  290 ACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGAT 339
      |||   |||   |||      :::::::::|||::::::   |||:::
 1861 oArgGluGln...ProArgProMetGluArgProAlaMetArgPheLys. 1876

  340 TCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCAT... 386
                           :::         |||||||||   |||
 1877 ...................GlnGluProGluGlnProLeuHisHisAsp 1886

  387 CCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCAC 436
      |||:::   ::::::      |||:::|||            ||||||
 1887 ProPheGlnAsnPheMetProProGlnArgSerMetGlnProArgAlaGl 1903

  437 AGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTAC 486
       :::      ::::::|||::::::
 1903 uAlaAlaProMetGlyArgGlnAsp........................ 1911

  487 GACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCG 536
                                  ||||||:::       :::   :
 1912 .........................ProProArgSer..MetAlaProS 1919

  537 CAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGC 586
      ::||||||::::::      |||   ::::::::::::|||:::
 1919 erAlaProGlnSerTyrThrProProIleGlnAlaGlnProValArgAsn 1935

  587 TGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCC 636
           :::|||   :::      :::::::::|||||||:::   :::||
 1936 LeuLeuSerGluSerValProProGlnArgThrProProGlnLeuSerPr 1952

  637 GAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGA 686
```



seq_name: pir2:E86185

seq_documentation_block:
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: E86185
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: E86185
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-402 <STO>
A;Cross-references: GB:AE005172; NID:g2388584; PIDN:AAB71465.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

alignment_scores:
Quality: 135.00 Length: 309
Ratio: 0.849 Gaps: 9
Percent Similarity: 51.456 Percent Identity: 24.595

alignment_block:
US-09-303-518D-463 x E86185 ..

```
Align seg 1/1  to: E86185  from: 1  to: 402

 86 ACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCCGAC 135
    |||        |||:::      ::::::::::::::::|||:::   ||
 68 ThrAlaGluSerSerSerSerThrLysSerAlaThrMetSerGlySerTh 84

136 GGGAAATACCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGG 185
    |   :::|||||   ||||||:::::::::|||      |||:::::::
 84 rThrHisThrThrSerSerAlaThrAlaSerSerThrAlaSerThrSerT 101

186 CCATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGA 235
    ::   |||         :::|||:::::   |||      |||
101 hrSerSerTyrSerThrSerTyrSerThrSerSerThrLysThr..... 115

236 TTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCC 285
       :::          :::   ||||||:::|||   ||||||   ||
116 ...ThrThrMetThrGlySerThrIleSerThrThrAlaSerAlaAlaPr 131

286 GATCACG.....GGCA 296
    |   :::      :::|
131 oThrSerThrAlaSerThrSerThrSerSerTyrSerThrSerTyrSerT 148

297 CAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACG 346
    ||:::|||   :::   :::|||:::                  |||
148 hrSerSerThrLysThrThrThrValThrGlySerThrIleGlyThrThr 164

347 AAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGAC 396
          ::::::|||   :::   |||:::   |||||||:::   :::::
165 AlaSerAlaAlaProThrSerThrSerThrSerThrAlaAsnSerSerAl 181

397 GGATACGAACACCATCCCG.....CCGACGGCTATGACGG 431
    ::::::::::|||   |||               ||||||||||||||||
181 aSerSerThrThrAsnProSerSerGlySerLysProThrAlaMetThrG 198

432 GCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACA 481
    ||      |||   :::   |||         :::   :::      |||
198 lyThrThrAlaAsnThrSerProSerAlaProThrSerSerProSerThr 214

482 GCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGAC 531
    :::::::::   ::::::          |||:::::::::   |||||
215 ThrAsnSerSerSerThrAlaAlaTyrThrSerSerGlySerLysProTh 231

532 AACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGC 581
    ||||::::::                  |||:::::|||      :::
231 rThrValThr.....ArgThrThrAlaAsnThrSerSerSerA 244

582 TATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACA 631
                      :::|||:::|||   :::|||   ::::::
244 la.....SerThrSerSerAlaSerProThrAsnSer 254

632 GCCCCGAGCTGG...ACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGC 678
    :::   |||      |||:::   :::        ||||||:::|||
255 SerThrSerThrProThrAsnSerSerAlaGlySerLysProThrThrMe 271

679 ACTGCAGATATCGTCAAAAACATCATCG.....G 707
              :::::::::|||||||||      :
271 tThrGlyThrThrThrAsnThrSerSerThrThrThrThrSerSerAlaS 288

708 CGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCG 757
    ::      ::::::   |||:::::|||                  :::
288 erThrThrLysSerSerSerSerSerAlaThrAsnSerSerSerGlySer 304

758 AAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAA 807
    |||   :::|||||||   :::   ||||||:::::::::       ::
305 LysProSerThrLeuSerThrThrThrAlaTyrThrAlaThrThrSerSe 321

808 AACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGA 857
    :         ::   ||||||                  :::::::
321 rProThrAlaGluProSerThrThrThrAlaSerLysProAlaThrSerS 338

858 CTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCG 907
    ::   |||      ||| |||   :::         ||||||:::   ||
338 erThrProProAlaPro.ProThrIleIleValSerLysProArg...Ar 353

908 CACAAGGCATAGAAGCCGTCAGCAA 932
    ||||:::|||||||||      |||
353 gThrHisHisArgSerIleProGln 361

seq_name: pir2:T15087

seq_documentation_block:
hypothetical protein K06A5.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15087
R;Wamsley, P.
submitted to the EMBL Data Library, December 1997
A;Description: The sequence of C. elegans cosmid K06A5.
A;Reference number: Z18291
A;Accession: T15087
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-919 <WAM>
A;Cross-references: EMBL:AF039038; NID:g2736359; PID:g2736360; PIDN:AAB94168.1; GSPDB
A;Experimental source: strain Bristol N2; clone K06A5
C;Genetics:
A;Gene: CESP:K06A5.4
A;Map position: 1
A;Introns: 14/3; 77/3; 99/3; 287/3; 448/3; 493/3; 611/3; 819/2; 868/3


alignment_scores:
           Quality:  135.00            Length:    696
             Ratio:  0.456              Gaps:     32
 Percent Similarity:  42.529   Percent Identity: 20.546

alignment_block:
 US-09-303-518D-463 x T15087   ..

 Align seg 1/1  to: T15087  from: 1  to: 919

138 GAAATACCACCTATTCGGCAGCAGGGG..... 164
    |||      |||:::|||:::|||||
283 GluGlnThrProValArgLysGlnGlyProIleAlaSerSerThrProGl 299

165 .....GGAGCTTGCCNAGCGCAACGGCCATATC 192
                    ||||||          |||   :
299 nGlnLysGlnArgLeuAlaAspGlyAlaAsnAsnGlnIleProProThrG 316

193 GGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGATTCAACA 242
    ::   ::::::   :::      :::|||   ||||||
316 lnLysSerGlnAspSerValGlnAlaValGlnProProProProArgPro 332

243 GGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACG 292
    ::::::|||      ::::::      :::   ||||||
333 AlaAlaArgAsnAlaGlnPheAlaSerAspAlaAspLeuPheAlaValPr 349

293 GGCACAAATTCCA..... 305
                |||
349 oLysAlaProProSerLysSerValArgAsnLeuAlaAlaSerAsnValA 366

306 .....TTCGCCCTTCGACAACCATGCCTC...ACATTCCGATTCTGACGA 347
         ||||||      :::      |||   ||||||:::|||
366 spIlePheAlaAspValAspSerValLeuAspThrPheHisPhe..... 380

348 AGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACG 397
                          :::      |||         |||||
381 .....GluSerThrPro.....GlyAr 386

398 GATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGC 447
    |:::|||   |||:::||||||::::::         :::   :::
386 gValArgLysProGlyArgArgAsnValSerSerProSerProGluProA 403

448 TATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAAAAGG 497
          ||||||:::   |||:::||||||   |||   |||:::::::::
403 rgHisArgSerSerSerArgAspGlyTyrGluGlnSerArgTyrSerGln 419
```

```
 498 CGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGAC 547
     |||         |||      :::   :::|||      :::|||:::::
 420 Arg.........TyrGluHisAspAsnSerArgTrpSerArgHisAsnAl 433

 548 AACGGCTTGCCGA..................................... 560
     :|||      |||
 433 aThrTyrArgArgHisGluAspGluSerArgMetSerArgLysArgSerI 450

 561 ...........CCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGT 599
                 |||       ||||||:::   |||:::   :::
 450 leValArgAspAspPheGluTyrSerArgArgHisAspAspGly...Ala 465

 600 AGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGAT 649
     |||||||||                 ||||||
 466 ArgArgArgAspTyrTyrAspAlaAspIleGln.................. 476

 650 CGGGCAATGCCGCCGA.......................AGCCTTCAAC 675
      |||:::   :::|||                      |||      |
 477 .GlyAspSerLysArgTyrArgGlyArgAspAlaSerSerSerSerGlyA 493

 676 GGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGC......AGGAGA 719
     ||      |||:::::::::::::|||:::|||:::|||       |||
 493 rgSerValArgPheGluGluGluHisArgArgHisGlyAspGluTyrArg 509

 720 AATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACA 769
     :::   |||      |||           |||   ||||||   :::||
 510 AspProArgGlyProArgAspTyrAsnAspTyrGlyArgArgArgAsnHi 526

 770 TTGCTGTCATGCACGGCTTGGG.................TCTGCTTTCC 801
     |   :::::::::|||   |||                 :::|||:::
 526 s...AlaAsnSerArgSerGlyGluAspGluGluLysLeuAsnAlaIleV 542

 802 ACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACT 851
       |||::::::        ::::::|||:::   :::       :::
 542 alArgArgGluLysGluLeuArgAsnArgLeuGlnLys.........Ser 555

 852 CAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCA 901
     |||:::          ||||||:::
 556 GlnLysAlaSerSerSerSerTyr........................... 563

 902 ATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATC 951
               ||||||||||||:::::::::            ::::::
 564 ..........ArgHisArgSerAsnSerSer.........AspAlaGluG 574

 952 CCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCGGCAT 1001
           ::::::|||::: ::::::::::::
 574 luSerLeuAsnGluTrpAsp.IleGluAsnGln.................. 584

1002 CACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCCGA 1051
           :::   :::   :::|||:::|||   :::   ::::::||||
 585 .......GluLeuLeuAspAsnSerMetMetPheGlyAspGlyIleProL 599

1052 AAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAAA 1101
     ||   ::::::|||   :::|||:::|||:::
 599 ysArgSerAsnAlaArgLysAspLysPheVal.................... 609

1102 TACCCGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCG 1151
                                  ::::::          ||
 610 ..............................LysLysGlnAlaThrAr 615

1152 TTACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCA 1201
     |           |||   |||:::|||   :::         :::   |
 615 gSerLysProAlaAsnSerThrLysSerProAlaGlnAlaArgLysLysL 632

1202 AAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTT 1251
     ||:::::::::|||   |||:::|||      :::   ::::::
 632 ysArgAlaSerLeuGluAspAsnArgAspLeuAsnAspSerIleAlaCys 648

1252 GACGGTAAAGGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACG.. 1299
     :::                        :::|||      |||
 649 AsnArgProArgArgSerCysValThrProValAlaLysLysIleThrTr 665

1300 ....AAGCTCGATATTCAAGAATTATCGGGGGGCGGTATACCTAAGGCTA 1345
          |||   |||:::::::::|||:::               |||
  665 pArgLysGlnAspLeuAspArgLeuLysArgValIleAlaLeuLysLysP 682

1346 AGCCTGTGTTTGATGCGAAACCGAGATGG...GAGGTTGATAGGAAGCTT 1392
          :::   ||||||         |||   ||||||   |||   |||
  682 roSerAlaSerAspAlaAsp......TrpThrGluValLeuArgLeuLeu 696

1393 AATAAA.................................TTGACAAC 1406
          |||                                   ||
  697 AlaLysGluGlyValValGluProGluValValArgGlnIleAlaIleTh 713

1407 TCGT.........................GAGCAGGTGGAGAAAAATG 1429
          ||||                   |||:::|||   |||:::|
  713 rArgLeuLysTrpValGluProGluGlnAsnGluGluValLeuLysGlnV 730

1430 TTCAGGAAACG.........AGAAGAAGGAGTCAGAGTAGTCAGTTTAAA 1470
     ||:::|||:::         |||||||||:::      ::::::   |||
  730 alGluGluValGluGlnLysArgArgArgGlyAlaValAlaArgValLys 746

1471 GCCCATGCGCAACGAGAATGGGAAAATAAAACAGGGTTAGATTTTAATCA 1520
        :::::::::   :::   |||   :::
  747 GluAsnValLysMetHisGluGluLeuArgGlu.................. 757

1521 TTTTATAGGTGGTGATATCAATAAGAAAGGCACAGTAACAGGAGGGCATA 1570
             ||||||:::   :::   :::          :::|||   :::|
  758 .......GlyGlyAsnHisArgAlaGluAspLeuGlnSerGlyValGluS 772

1571 GTCTAACCCGTGGTGATGTACGGGTGATACAACAAACCTCGGCACCTGAT 1620
     ||:::                              :::         |||:::
  772 erMet..........................GluAspTyrGlnProGlu 779

1621 AAACATGGGGTTTATCAAGCGACAGTGGAAATTAAAAAGCCTGATGGAAG 1670
             ::::::   |||:::   :::   ::::::   |||
  780 AspValAlaAlaAspGlnSerLeuLeuAlaLeuArgThrPro........ 793

1671 TTGGGAGGTGAAAACGAAAAAAGGTGGGAAAGTGATGACCAAGCACACCA 1720
             :::   |||||||||   |||         |||:::   ::::
  794 .......IleValThrLysLysArgGlyGly......ThrArgAlaSerI 806

1721 TGTTCCCAAAAGATTGGGATGAGGCTAGAATT.................. 1752
     :::::||||||      :::::::::   :::
  806 leMetProLysProValGluAspSerProMetSerArgGlyAsnAsnSer 822

1753 ..............AGGGCTGAAGTTACTTCGGCTTGGGAA........ 1779
                     |||   |||   |||:::|||   :::
  823 ThrPheAsnSerProArgLeuGluGlnThrLysAlaLysAspIleGluTh 839

1780 .............................................A 1780
                                                 |
  839 rAsnPheLysTyrValGlnHisLeuSerMetMetGlnAlaArgProSerS 856

1781 GTAGAATA............ATGCTTAAGGATAATAAATGGCAGGGTACA 1818
     |||||:::               ::::::::::::   ::::::|||:::
  856 erArgLeuLysLysSerSerSerMetAsnAsnSerThrTyrArgGlyAsn 872

1819 AGTAAATCGGGTATTAAAATAGAAGGATTTACCGAA 1854
     ::::::::::::|||::::::|||      |||:::
  873 LysAsnThrSerIleSerLeuGluLysGlyThrGln 884
```

seq_name: pir2:T14342

seq_documentation_block:
NSD1 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14342
R;Huang, N.; vom Baur, E.; Garnier, J.M.; Lerouge, T.; Vonesch, J.L.; Lutz, Y.; Chamb
EMBO J. 17, 3398-3412, 1998
A;Title: Two distinct nuclear receptor interaction domains in NSD1, a novel SET prote
A;Reference number: Z17993; MUID:98292459
A;Accession: T14342
A;Status: preliminary; translated from GB/EMBL/DDBJ

```
A;Molecule type: mRNA
A;Residues: 1-2588 <HUA>
A;Cross-references: EMBL:AF064553; NID:g3329464; PID:g3329465; PIDN:AAC40182.1


alignment_scores:
           Quality:  135.00            Length:    468
             Ratio:  0.640              Gaps:     25
 Percent Similarity:  45.085   Percent Identity: 23.718

alignment_block:
 US-09-303-518D-463 x T14342   ..

 Align seg 1/1  to: T14342  from: 1  to: 2588

      83 CAAACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCC 132
         |||   :::|||||||||   :::      |||   |||   :::::::||
    2158 GlnLeuLeuProLeuSerLysLysAlaLeuThrGlySerCysGlnArgPr 2174

     133 G.........................ACGGGAAATACCACCTAT.... 151
         |                       |||   ::::::::::::
    2174 oLeuLeuProGluArgProProGluArgThrAspSerSerSerHisLeuL 2191

     152 .............TCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCC 187
                     :::   ||||||::::::      |||   :::
    2191 euAspArgIleArgAspLeuAlaGlySerGlyThrLysSerGlnSerLeu 2207

     188 ATATCGGATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGATT 237
         :::|||:::   :::
    2208 ValSerSerGlnArgProGln........................... 2214

     238 CAACAGGCGGCCGTTGAAGGAA.........ATATCGGCTACATTGTCCG 278
          ::|||   |||   ||||||         :::          |
    2215 .AspArgProProAlaLysGluGlyProArgProGlnProProAspArgA 2231

     279 CTTTTCCGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCCT 328
         ||   |||:::|||      :::|||           ::::::::::|||
    2231 laSerProMetThrArgProSerSerSerProSerValSerSerLeuPro 2247

     329 CACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTT 378
                 |||   :::|||   :::   :::|||||||||   ::
    2248 ............LeuGluArgPro...LeuArgMetThrAspSerArgLe 2259

     379 TACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGA 428
         :      |||||||||
    2259 uAspLysSerIleGly............................... 2264

     429 CGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATAT 478
                 ||||||:::         ||||||:::
    2265 ...........AlaAlaSer.........ProLysSer............ 2270

     479 ACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACC 528
                     :::|||:::   |||      |||||||||
    2271 ..............GlnAlaValGluLysThrProAlaSerThrGlyLe. 2282

     529 GACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGG 578
             ::::::||||||:::   |||   |||   |||      |||
    2282 uArgLeuSerSerProAspArgLeuLeuThrThrAsnSerProLysProG 2299

     579 CGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGAT 628
           :::                          |||:::   |||||||:::
    2299 lnIle.........................SerAspArgProProGlu 2306

     629 ACAGCC...CCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTTCAAC 675
            :::      |||   |||:::|||      ||||||:::
    2307 LysSerHisAlaSerLeuThrGlnArgLeuProProProGluLysValLe 2323

     676 GGCACTGCAGATATCGTCAAAAACATCATCGGCGCGG............. 712
          :::      ::::::|||   :::::::::   :::
    2323 uSerAlaValValGlnSerLeuValAlaLysGluLysAlaLeuArgProV 2340

     713 ..CAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAG 760
           :::::::::   :::::::::      |||            |||

    2340 alAspGlnAsnThrGlnSerLysHisArgPro.....AlaVal 2352

     761 GCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAAC 810
         ::::::   |||::::::   |||                   |||:::
    2353 ValMetAspLeuIleAspLeuThr.................ProArgGl 2363

     811 AAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTA 860
          :::   ||||||::::::
    2363 nLysGluArgAlaAlaSer............................. 2369

     861 TGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCAC 910
           |||||||:::   :::      :::::::::||||||   |||   :::
    2370 ..ProGlnGluValThrProGlnAlaAspGluLysThrAlaMetLeuGlu 2385

     911 AAGGCATAGAAGCCGTCAGCAATATCTTTATGG................. 943
         ::::::      |||||||:::   :::|||
    2386 SerSerSerTrpProSerSerLysGlyLeuGlyHisIleProArgAlaTh 2402

     944 ....CAGCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGG 989
             :::   |||   |||:::::::|||:::   |||||||:::   :::|
    2402 rGluLysIleSerValSerGluSerLeuGlnProSerGlyLysValAlaA 2419

     990 CT..............TGGGCGGCATCACGGCACATCCTGTCAAGCGGT 1024
         ||              |||   |||   :::   :::
    2419 laProSerGluHisProTrpGlnAlaValLysSerLeu............ 2431

    1025 CGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGAC 1074
                        :::|||:::|||      :::||||||||||||
    2432 ..............ThrHisAlaArgPheLeuSerProProSerAlaLy 2443

    1075 AATTTTGCCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCG 1124
             :::                                      :::   |
    2443 sAlaPhe.....................................LeuTyrG 2448

    1125 AAATATCCGTTCAAACT......TGGAGCAGCGTTACGGCAAAGAAAACA 1168
         ||   ::::::|||:::          :::   |||   |||:::::::|||
    2448 luSerAlaThrGlnAlaSerGlyArgThrProValGlyAlaGluGlnThr 2464

    1169 ...TCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTCAAA... 1212
            :::||||||:::|||:::  |||:::      |||:::||||||
    2465 ProGlyProProSerProAla..ProGlyLeuValLysGlnValLysGln 2480

    1213 CTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAAAGG 1262
         |||:::                |||:::|||      |||   :::   ||
    2481 LeuSerArgGlyLeuThrAlaLysSerGlyGlnSerPheArgSerLeuGl 2497

    1263 G..............TTTCCGAATTTTGAGAAGCACGTGAAATATGATA 1297
         |               :::|||||   ||||||
    2497 yLysIleSerAlaSerLeuProAsnGluGluLys............... 2508

    1298 CGAAGCTCGATATTCAAGAATTATCGGGGGGCGGTATACCTAAGGCTAAG 1347
           ||||||         |||   |||      |||:::   ||||||:::
    2509 ..LysLeuThrThrThrGluGlnSerProTrpGlyLeuGlyLysAlaSer 2524

    1348 CCT 1350
         |||
    2525 Pro 2525

seq_name: pir2:T34369

seq_documentation_block:
hypothetical protein T19D12.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34369
R;Favello, A.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid T19D12.
A;Reference number: Z21513
A;Accession: T34369
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1777 <FAV>
```

```
A;Cross-references: EMBL:U41263; PIDN:AAC24428.1; GSPDB:GN00020; CESP:T19D12.1
A;Experimental source: strain Bristol N2; clone T19D12
C;Genetics:
A;Gene: CESP:T19D12.1
A;Map position: 2
A;Introns: 36/1; 134/2; 180/1; 622/3; 691/2; 754/1; 1111/2; 1174/1; 1271/3; 1322/2; 1681

alignment_scores:
            Quality:  134.50            Length:    505
              Ratio:  0.570              Gaps:     25
 Percent Similarity:  46.733   Percent Identity: 23.564

alignment_block:
 US-09-303-518D-463 x T34369   ..

 Align seg 1/1  to: T34369  from: 1  to: 1777

     44 CAGTGTGCCTGCCGATGCATGCACACGCCTCAGATTTGGCAAACGATCCC 93
        |||     |||          :::|||        :::|||:::||
    191 GlnSerAspCysProSerThrValAlaProValThrCysSerThrMetPr 207

     94 TTTATCCGGCAGGTTCTCGACCGTCAGCATTTCGAACCCGACGGGAAATA 143
        |:::                |||:::   :::   |||      :::|
    207 oMetProThrThrGlnProLeuValThrSerThrAlaProGluValThrT 224

    144 CCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCATATCG 193
        |||||  :::   :::      :::        :::||||||  :::
    224 hrThrValAlaGlnThrThrThrAlaProIleValThrThrAlaAsnThr 240

    194 GATTGGGAAACATACAAAGCCATCAGTTGGGCCACCTGATGATTCAACAG 243
           :::      :::   :::   :::|||         :::||
    241 ThrThrGlnGlyValThrThrThrAlaGlyValThrThrThrValThrAr 257

    244 GCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTCCGATCACGG 293
        |         :::   ::::::||||||   :::|||   |||   |||:
    257 gAlaGlnAsnSerThrLeuAlaAlaThrThrThrAla...ProSerThrA 273

    294 GCACAAATTCCATTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTG 343
        ::|||::::::         :::||||||:::
    273 snThrThrThrGlnGlyValThrThrThrValGlyLys..... 285

    344 ACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGG 393
        |||      |||:::        :::|||:::::::::|||:::
    286 ThrThrThrValThrThrAlaGlnAsnSerThrTrpAlaAlaThrThrTh 302

    394 GACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGG 443
         :::::::::::::|||      |||::::::   |||:::
    302 rAlaSerAsnThrThrThrGlnProValValThrThrSer..... 315

    444 CGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATAA 493
                      :::::::::||||||   |||   ||||||
    316 .............ThrSerThrGlnGlyIleSerThr...ThrThrAla 326

    494 AAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCA..... 538
        :::|||   |||      |||:::|||:::   ||||||||||:::
    327 GlnAlaThrPro......SerSerSerValIleProThrThrThrGlnTh 341

    539 .........GCACCGGACAACGGCTTGCCG...ACCGTTTCCACAATGC 575
                      ::::::|||:::|||   ||||||||||||
    341 rThrGlnArgProThrSerThrGlyIleProSerThrValSerThrSerG 358

    576 CGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCCA... 622
                                  :::|||:::|||::::::|||
    358 ln.......................GlyThrSerSerThrSerProIle 366

    623 CCCGATACA.:.................GCCCCGAGCTGGACAGATCGGGC 654
        |||:::|||                  |||||||||
    367 ProSerThrThrGlnThrSerSerSerAlaProSerThrTyrThrSerAs 383

    655 AATGCCGCCGAAGCCTTCAACGGCACTG..................... 682
         :::   |||:::|||:::|||:::|||
    383 nPheThrProSerProThrThrThrLeuLeuThrSerThrIleAlaProS 400

    683 .............CAGATATCGTCAAAAACATCATCGGCGCGGCAG... 715
                         |||||||||:::||||||   ::::::
    400 erThrGlnGlyValProThrSerSerLysSerSerSerProAsnSerThr 416

    716 ......GAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAA 759
                :::::      |||          :::   :::::
    417 ProThrThrThrIleThrProGlyAlaProSerSerThrLeuGlySerSe 433

    760 GGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAA 809
        :::::::|||:::::::|||   |||            :::   |||:::|
    433 rSerSerThrIleValSerThrThr.............IleThrProSerT 446

    810 CAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACT 859
        ||      ::::::||||||:::                          |||
    446 hrPro...LysValSerThrLeu.......................Thr 453

    860 ATGCCGCAGCAGCCA................................TC 876
        :::   |||:::|||                                ||
    454 LeuSerGlnSerProThrProThrSerThrProLeuValValSerSerSe 470

    877 CGCGATTGGGCAGTCCAAAACCCCAATG..............CCGCACA 911
        |:::   |||:::|||:::|||   :::               |||
    470 rSerSerGlySerSerSerThrValValThrSerThrIleThrProSerT 487

    912 AGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAG 961
                  |||::::::   |||      ||||||:::||||||
    487 hrGlnGlyValProThrSerThrSerAsn...GlnProThrProSerThr 502

    962 GGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCG.............. 997
        :::         :::   ::::::|||:::   |||
    503 SerAsnProThrThrProLysSerThrValThrAlaSerProSerThrTh 519

    998 ...............GCATCACGGCACATCCTGTCAAGCGGTCGCAGAT 1031
                        ||||||         ::::::|||:::
    519 rGlyAlaThrSerThrAlaSerProSerThrIleThrSerSerAlaProT 536

   1032 GGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTG 1081
          ::::::||||||                  ||||||:::|||:::
    536 hrSerGlnSerHis..............SerProSerSerThrMetThr 547

   1082 CCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATC 1131
                                          |||:::|||      ||
    548 Ser............................ThrValProValThrSe 554

   1132 CGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAAC 1181
        |:::   :::   ::::::::::|||:::      |||      |||
    554 rThrPheAlaSerThrThrThrThrValProIleThrValAlaProGlyG 571

   1182 CGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACC 1231
          |||                              |||:::|||::::::
    571 lnCys.......................TyrCysGlnSerAsnValAla 579

   1232 CGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAATTTTGAGAAG 1281
              |||::::::||||||                  :::      ::
    580 ......ValAlaPheGluLeuThrSerGlyThrSerAspLeuAspLeuAs 594

   1282 CACGTGAAATATGATACGAAGCTCGATATTCAAGAATTATCGGGGGGCGG 1331
        :      |||:::|||   :::   :::|||      |||   ||||||
    594 pIleGlnAsnPheIleAlaAsnTyrLeuPhe...PheTyrSerGlyAlaP 610

   1332 TATACCTAAGGCTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAGGTTG 1381
          |||:::      :::|||                  |||:::|||
    610 roTyrIleLeuGlyAspLeu..............ThrAspAsnArgThr 621

   1382 ATAGGAAGCTTAATA 1396
              ||||||:::
    622 AlaIleSerLeuVal 626

seq_name: pir2:I54632

seq_documentation_block:
tsh protein - Escherichia coli
```

```
C;Species: Escherichia coli
C;Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 08-Dec-2000
C;Accession: I54632
R;Provence, D.L.; Curtiss, R.
Infect. Immun. 62, 1369-1380, 1994
A;Title: Isolation and characterization of a gene involved in hemagglutination by an avi
A;Reference number: I54632; MUID:94178945
A;Accession: I54632
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1377 <RES>
A;Cross-references: GB:L27423; NID:g469235; PIDN:AAA24698.1; PID:g469236
C;Superfamily: IgA-specific metalloendopeptidase


alignment_scores:
           Quality:  134.00            Length:    707
             Ratio:  0.428              Gaps:     36
 Percent Similarity:  44.272   Percent Identity: 20.509

alignment_block:
 US-09-303-518D-463 x I54632   ..

 Align seg 1/1  to: I54632  from: 1  to: 1377

     109 CTCGACCGTCAGCATTTCGAA.............CCCGACGGGAAATA 143
         ||||||:::::::::::::                 :::|||
     182 LeuAspLysGluArgTyrProValPheTyrArgLeuGlySerGlyThrGl 198

     144 CCACCTATTCGGCAGCAGGGGGGAGCTTGCCNAGCGCAACGGCCATATCG 193
          ::::::      |||:::|||:::|||:::      :::|||      :
     198 nTyrIleLysAspSerAsnGlyGlnLeuThrGlnMetGlyGlyAlaTyrS 215

     194 GA.................TTGGGAAACATACAAAGCCATCAGTTGGGC 225
         ::                  :::|||:::::::::|||:::|||   |||
     215 erTrpLeuThrGlyGlyThrValGlySerLeuSerSerTyrGlnAsnGly 231

     226 CACCTGATG..................................... 234
         :::::::::
     232 GluMetIleSerThrSerSerGlyLeuValPheAspTyrLysLeuAsnGl 248

     234 ..................................................... 234

     248 yAlaMetProIleTyrGlyGluAlaGlyAspSerGlySerProLeuPheA 265

     235 ...........ATTCAACAG.............................. 243
                     :::|||:::
     265 laPheAspThrValGlnAsnLysTrpValLeuValGlyValLeuThrAla 281

     244 .........GCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGCTTTTC 284
                 |||:::   :::||||||         ::::::   :::::
     282 GlyAsnGlyAlaGlyGlyArgGlyAsnAsnTrpAlaValIleProLeuAs 298

     285 CGATCACGGGCACAAATTCCATTCGCCCTTCGACAACCATGCC....... 327
         :      |||:::||||||:::::      ||||||   |||
     298 pPheIleGlyGlnLysPheAsnGlu......AspAsnAspAlaProValT 313

     328 ................................TCACATTCCGATTCT 342
                                        |||   :::::::|||
     313 hrPheArgThrSerGluGlyGlyAlaLeuGluTrpSerPheAsnSerSer 329

     343 GACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTG 392
               ||||||:::   :::::::|||   :::   |||   :::|||
     330 ThrGlyAlaGlyAlaLeuThrGlnGlyThrThrThrTyrAlaMetHis.. 345

     393 GGACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCG 442
                                                |||   ||||||:
     346 ................................GlyGlnGlnGlyA 350

     443 GCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGACATA 492
         ::         |||   |||:::         |||::::::
     350 snAspLeuAsnAlaGlyLysAsnLeu......IlePheGln......... 361

     493 AAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCAC 542
                 :::||||||   ::::::||||||   |||:::   |||
     362 .........GlyGlnAsnGlyGlnIleAsnLeuLysAspSerValSerGl 375

     543 CGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGC 592
          |||                          ||||||::::::      :
     375 nGly.........................AlaGlySerLeuThrPheA 383

     593 AAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTG 642
         ::          |||::::::      :::|||
     383 rg.........AspAsnTyrThrValThrThr..................... 390

     643 GACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGT 692
                     :::   :::::::::::::|||::::::   :::||
     391 ............SerAsnGlySerThrTrpThrGlyAlaGlyIleValVa 403

     693 CAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCA...GGCGATG 739
         |   |||   :::::::::      ::::::   |||:::   ||||||
     403 lAspAsnGlyValSerValAsnTrpGlnValAsnGlyValLysGlyAspA 420

     740 CCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTG 789
           ::::::   |||:::|||||||:::         :::::::::|||
     420 snLeuHisLysIleGlyGluGlyThr......LeuThrValGlnGlyThr 434

     790 GGTCTG............CTTTCCACCGAAAACAAGATGGCGCGCATCAA 827
         |||:::            |||:::::::   :::      :::   :::||
     435 GlyIleAsnGluGlyGlyLeuLysValGlyAspGlyLysValValLeuAs 451

     828 CGATTTGGCAGAT...ATGGCGCAACTCAAAGACTATGCCGCAGCAGCCA 874
         |:::   ||||||      :::|||::::::   ::::::::::::   |
     451 nGlnGlnAlaAspAsnLysGlyGlnValGlnAlaPheSerSerValAsnI 468

     875 TCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCC 924
         ||          |||      :::|||:::::::::         |||
     468 le.........AlaSerGlyArgProThrValValLeuThrAspGluArg 481

     925 GTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGCTGT 974
               |||          ::::::          :::|||
     482 GlnValAsnProAspThrValSerTrpGlyTyrArgGly............ 494

     975 CCGGGGAAAATACGGCTTGGGCGGC......ATCACGGCACATCCTGTCA 1018
             |||          :::::::|||      :::|||   |||   :::|
     495 ....GlyThrLeuAspValAsnGlyAsnSerLeuThrPheHisGlnLeuL 510

    1019 AGCGGTCGCAGATGGGCGCGATC.......................... 1041
         ||   ::::::   ||||||:::
     510 ysAlaAlaAspTyrGlyAlaValLeuAlaAsnAsnValAspLysArgAla 526

    1042 .................GCATTGCCGAAAGGGAAATCCGCCGTCAGC.. 1071
                           ||||||          |||   |||::::::
     527 ThrIleThrLeuAspTyrAlaLeuArgAlaAspLysValAlaLeuAsnGl 543

    1072 ................GACAATTTTGCCGATGCGGCATACGCCAAATACC 1105
                          :::   |||          |||   ||||||
     543 yTrpSerGluSerGlyLysGlyThrAlaGlyAsnLeuTyr...LysTyrA 559

    1106 CGTCCCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAG...CGT 1152
           :::||||||   :::   :::          |||:::|||
     559 snAsnProTyrThrAsnThrThrAspTyrPheIleLeuLysGlnSerThr 575

    1153 TACGGC..................AAAGAAAACATCACCTCCTCAACCGT 1184
         ||||||                 ::::::|||   |||   :::   ||
     576 TyrGlyTyrProThrAspGlnSerSerAsnAlaThrTrpGluPheVa 592

    1185 GCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACCCGA 1234
         |      |||:::|||          ||||||:::   :::|||
     592 lGlyHisSerGlnGlyAspAlaGlnLysLeuValAlaAspArgPheAsnT 609

    1235 AGACAGGCGTACCGTTTGACGGT......AAAGGGTTTCCGAATTTTGAG 1278
           :::|||      |||   |||      ||||||      |||   :::
     609 hrAlaGlyTyrLeuPheHisGlyGlnLeuLysGlyAsnLeuAsnValAsp 625

    1279 AAGCACGTGAAA.....................TATGATACGAAGCT 1304
```

```
      :::::::::                                  |||
  626 AsnArgLeuProGluGlyValThrGlyAlaLeuValMetAspGlyAlaAl 642

 1305 CGATATTCAA...........GAATTATCGGGGGGCGGTATACCTAAGG 1342
       ||||||:::           :::   :::|||            ::::
  642 aAspIleSerGlyThrPheThrGlnGluAsnGlyArgLeuThrLeuGlnG 659

 1343 CTAAGCCTGTGTTTGATGCGAAACCGAGATGGGAGGTTGATAGGAAGCTT 1392
      ::   ||||||:::   |||
  659 lyHisProValIleHisAla........................... 665

 1393 AATAAATTGACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGAAACGAG 1442
                 :::   ::::::|||      ::::::     :::
  666 .........TyrAsnThrGlnSerValAlaAspLysLeuAlaAlaSerGl 679

 1443 AAGAAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGGG 1492
        :::|||      :::|||      :::   :::|||::::::||||
  679 yAspHisSerValLeuThrGlnProThrSerPheSerGlnGluAspTrpG 696

 1493 AAAATAAAACA.........................GGGTTAGATTTT 1515
      |||||::::::                        :::   ||||||
  696 luAsnArgSerPheThrPheAspArgLeuSerLeuLysAsnThrAspPhe 712

 1516 AATCATTTTATAGGTGGTGATATCAATAAGAAAGGCACAGTAACAGGAGG 1565
      :::      :::|||   :::   :::   :::   |||:::   :::
  713 Gly......LeuGlyArgAsnAlaThrLeuAsnThrThrIleGlnAlaAs 727

 1566 GCAT...AGTCTAACCCGTGGTGATGTACGGGTGATACAACAAACCTCGG 1612
       :::   |||:::|||   ||||||   |||||||:::
  727 pAsnSerSerValThrLeuGlyAspSerArgValPhe.............. 739

 1613 CACCTGATAAACATGGGGTTTATCAAGCGACAGTGGAAATTAAAAAGCCT 1662
                                          |||   |||
  740 .....................................IleAspLysAsn 743

 1663 GATGGAAGTTGGGAGGTGAAAACG...AAAAAAGGTGGGAAAGTGATGAC 1709
      ||||||:::      :::   |||   :::::::|||   :::|||   ||
  744 AspGlyGlnGlyThrAlaPheThrLeuGluGluGlyThrSerValAlaTh 760

 1710 CAAGCACACCATGTTCCCAAAAGATTGGGATGAGGCTAGAATTAGGGCTG 1759
      |                  ||||||   |||::::::   :::::::::
  760 r..................LysAspAlaAspLysSerValPheAsnGlyT 771

 1760 AAGTTACTTCGGCTTGGGAAAGTAGAATAATGCTTAAGGATAATA...AA 1806
        |||:::         :::|||   :::   ::::::||| |||   ||
  771 hrValAsnLeuAspAsnGlnSerValLeuAsnIleAsnAsp.IlePheAs 787

 1807 TGGCAGGGTACAAGTAAAT 1825
      ||||   :::|||:::|||
  787 nGlyGlyIleGlnAlaAsn 793
```

seq_name: pir2:T38236

seq_documentation_block:
hypothetical protein SPAC23A1.17 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T38236
R;Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21780
A;Accession: T38236
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1611 <MUR>
A;Cross-references: EMBL:AL021813; PIDN:CAA16991.1; GSPDB:GN00066; SPDB:SPAC23A1.17
A;Experimental source: strain 972h-; cosmid c23A1
C;Genetics:
A;Gene: SPDB:SPAC23A1.17
A;Map position: 1

alignment_scores:

```
            Quality:  134.00              Length:    589
              Ratio:  0.563                 Gaps:     26
  Percent Similarity:  40.407   Percent Identity: 20.034
```

alignment_block:
US-09-303-518D-463 x T38236 ..

Align seg 1/1 to: T38236 from: 1 to: 1611

```
  308 CGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAGT 357
      :::|||   :::|||   |||      |||:::         |||   ||
  998 LysProValSerThrSerProAlaAlaProLeuAlaArgValProProVa 1014

  358 CCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATACGAACA 407
      ||||      :::|||
 1014 lProLysLeuSerSer.............................L 1020

  408 CCATCCCGCCGACGGCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTC 457
          ||||||:::   :::   :::      |||      ||||||:::
 1020 ysAlaProProValProLeuProSerAlaAspAlaProProIleProVal 1036

  458 CCAAAG...GCGCGAGGGATATATACAGCTACGACATAAAAGGCGTTGCC 504
      |||:::   |||      :::         |||:::         :::||
 1037 ProSerThrAlaProProValProIleProThrSerThrProProValPr 1053

  505 CAAAATATCCGCC.........TCAACCTGACCGACAACCGCAGCACCGG 545
      ||||   |||:::         |||:::   |||   :::   ||||||:
 1053 oLysSerSerSerGlyAlaProSerAlaProProProValProAlaProS 1070

  546 ACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAG 595
      :::::   :::|||::::::         ||||||
 1070 erSerGluIleProSerIle.........ProAla.............. 1078

  596 GAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGG.. 643
                     |||:::|||||||||      |||||||||
 1079 ............ProSerGlyAlaProProValProAlaProSerGlyIl 1091

  644 ...................ACAGATCGGGCAATGCCG...CCGAAGCCTT 671
                          :::      |||   |||   ||||||||||
 1091 eProProValProLysProSerValAlaAlaProProValProLysProS 1108

  672 CAACGGCACTG...CAGATATCGTCAAAAACATCATCGGCGCGGCAGGAG 718
      ||:::|||:::      :::   :::   :::::::::
 1108 erValAlaValProProValProAlaProSerGlyAlaProProValPro 1124

  719 AAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAAC 768
      |||   |||:::   |||   |||      |||   :::   |||
 1125 LysProSerValAlaAlaProProValProValProSerGlyAla..... 1139

  769 ATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGC 818
                                     ||||||
 1140 ..............................ProPro............. 1141

  819 GCGCATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAG 868
                           :::   :::   |||   :::   |||
 1142 ...................ValProLysProSerValAlaAlaProPro 1151

  869 CAGCCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATA 918
          |||:::      |||         :::|||
 1152 ValProAlaProSerGlyAlaProProValPro.................. 1162

  919 GAAGCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGG 968
       |||||||||:::   :::         |||:::||||||:::|||:::
 1163 .LysProSerValAlaAlaProProValProAlaProSerSerGlyIleP 1179

  969 AGCTGTCCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCTGTCA 1018

 1179 ro..................................................... 1179

 1019 AGCGGTCGCAGATGGGCGCGATCGCATTGCCGAAAGGGAAATCCG..... 1063
                                                    |||
 1180 ..........................................ProValPr 1182
```

```
1064 ....CCGTCAGCGACAATTTTGCCGATGCGGCATACGCCAAATACCCGTC 1109
1182 oLysProAlaAlaGlyValProProValProProProSerGluAlaProP 1199

1110 CCCTTACCATTCCCGAAATATCCGTTCAAACTTGGAGCAGCGTTACGGCA 1159
1199 roValProLysProSerValGlyVal..................... 1207

1160 AAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAAAAATGTC 1209
1208 ...........ProProValPro..ProProSerThrAlaProProVal 1219

1210 AAACTGGCAGACCAACGCCACCCGAAGACAGGCGTACCGTTTGACGGTAA 1259
1220 ProThrProSerAlaGlyLeuProProValProValPro...ThrAlaLy 1235

1260 AGGGTTTCCGAATTTTGAGAAGCACGTGAAATATGATACGAAGCTCGATA 1309
1235 sAlaProPro.......................ValProAlaProS 1243

1310 TTCAAGAATTATCGGGGGGCGGTATACCTAAGGCT.........AAGCCT 1350
1243 erSerGluAlaProSerValSerThrProArgSerSerValProSerPro 1259

1351 GTGTTTGATGCGAAACCGAGATGGGAGGTTGATAGGAAGCTTAATAAATT 1400
1260 HisSerAsnAlaSerProSerProThrSerSerSerMetAlaSerAlaAl 1276

1401 GACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGAAACGAGAAGAAGGA 1450
1276 aProAlaArgThrSerValSerArgSerLysSerLysAlaGluArgHisG 1293

1451 GTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGGGAAAATAAA 1500
1293 luThrSerThrSerSerArgLys.................SerSerLys 1303

1501 ACAGGGTTAGATTTTAATCATTTTATAGGTGGTGATATCAATAAGAAAGG 1550
1304 SerGlyGluHisHisHisHisHisAsnGluGlyHisAlaAspSerSerSe 1320

1551 CACAGTAACAGGA...GGGCATAGT..................... 1572
1320 rThrArgThrSerLeuAlaHisGlnAspSerArgLysSerLeuHisArgH 1337

1573 ..CTAACCCGTGGTGATGTACGGGTGATACAACAAACCTCGGCACCTGAT 1620
1337 isLeuSerArgSerSerSerArgAlaSerLysLysProSerIleValSer 1353

1621 AAACATGGGGTTTATCAAGCGACAGTGGAAATTAAA......AAGCCTGA 1664
1354 ThrThrGlyProPheAsnGluSerPheSerAlaLysProValGluProCy 1370

1665 TGGAAGT.........TGG........................... 1674
1370 sAlaSerGluLysTrpTrpLeuAsnSerThrAlaValProLysSerValV 1387

1675 ...............................GAGGTGAAAACGAAA 1689
1387 alGlnMetAsnAspSerValLeuTyrMetIleLysGluGlyIleThrGly 1403

1690 AAAGGTGGGAAAGTGATGACCAAGCACACCATGTTCCCAAAAGATTGGGA 1739
1404 GlnAspLysLysTyrLysSerValHisIleLeuPhePro...AspTyrSe 1419

1740 TGAGGCTAGAATTAGGGCT.....................:... 1758
1419 rGlnThrValLeuThrAlaThrPheAsnProHisAsnGlnAsnIleThrG 1436

1758 ..............................................  1758

1436 lnLeuSerGlnLeuGlnLeuAlaProProAlaGlnProSerLysAlaArg 1452

1759 ......GAAGTTACTTCGGCTTGGGAAAGTAGAATAATGCTTAAGGATAA 1802
```

```
          1453 LeuAspGluGluTyrAlaCysTyrGlySerThrIleLeuLysLysAlaAr 1469

          1803 TAAATGGCAGGGTACA 1818
          1469 gAlaTyrGlnGlySer 1474
```

seq_name: pir2:T18535

seq_documentation_block:
high molecular mass nuclear antigen - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18535
R;Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A;Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of ch
A;Reference number: Z18955; MUID:9803440
A;Accession: T18535
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1151 <SHI>
A;Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1

alignment_scores:
Quality: 133.50 Length: 486
Ratio: 0.678 Gaps: 23
Percent Similarity: 40.535 Percent Identity: 23.251

alignment_block:
US-09-303-518D-463 x T18535 ..

Align seg 1/1 to: T18535 from: 1 to: 1151

```
      86 ACGATCCCTTTATCCGGCAGGTTCTCGACC............GTCAGCAT 123
     177 ThrValProLysAlaThrAlaGluAlaThrAlaValThrAlaAlaSerGl 193

     124 TTCGAACCCG.........ACGGGAAATACCACCTATTCGGCAGCAGGGG 164
     193 nSerAlaProLysAlaAlaThrAspAlaAlaAlaValThrAlaAlaSerG 210

     165 GGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAAGCC 214
     210 lnSerAlaProLysAlaThrVal............GluValLysProAla 222

     215 ATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATC 264
     223 AlaAlaAlaValAla.......................LysGluAlaLy 231

     265 GGCT...ACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCATTCGCC 311
     231 sAlaValThrAlaAlaAlaAlaAlaProLysAlaThrAlaGluAlaLysP 248

     312 CTTCG...ACAACCATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTC 358
     248 roAlaProValThrSerProThrIleProCysSerSerAlaGluAlaLys 264

     359 CCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACG...GATACGAA 405
     265 ProLeuThrAlaAlaSerProThrAlaSerLysAlaThrAlaGluAlaLy 281

     406 CACCATCCCGCCGACGGCT.........ATGACGGGCCACAGGGCGGCGG 446
     281 sProValProAlaThrAlaSerLeuMetAlaThrLysValThrAlaGluA 298

     447 CTATCCCG..............CTCCCAAAGGCG.........CGAGGG 472
     298 laLysProAlaProSerProSerValProLysAlaThrThrAspThrLys 314

     473 ATATATACAGCTACGACATAAAAGGCG..................... 499
```

```
 315 AlaValThrAlaThrAlaProLysAlaGlyProAspValLysProAlaVa 331

 500 ..............................................T 500

 331 lAlaValCysAlaGluAlaLysProAlaProProProProProGlnGlnL 348

 501 TGCCCAAAATATCCGCCTCAACCTGACCGACA..................... 532
     |||||||   :::|||:::::   ||||||
 348 euProLysAlaAlaAlaAlaAlaAlaProThrGlyThrGluLeuLysPro 364

 533 ACCGCAGCACCGGACAACGGCTTGCCG..............ACCGTTTC 567
     :::::|||||||   :::|||   |||               ||||||::
 365 AlaThrAlaProProHisGlySerProArgAlaAsnSerHisThrValTh 381

 568 CACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACG.......... 607
     ::::   |||           |||         ||||||
 381 rValThrProProAsnValProArgAlaAlaAlaAlaThrValProThrA 398

 608 ...................GATTCAAACGCGCCACCCGATACAGCCCCG 637
                         ::::::::::::   |||   :::||||||
 398 laGlyAlaValProLysAlaSerThrGlyThrThrProAlaAlaAlaPro 414

 638 AGCTGG...ACAGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCA 684
     :::      :::   :::|||   |||      |||   :::      ||
 415 GlnGlnProValProLysAlaAlaProValThrProProSerProGlnGl 431

 685 GATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGG 734
     |         ::::::::::::|||                ::::|||   :
 431 nAlaValProArgAlaAlaThrAla.............AlaAlaAlaProV 444

 735 CGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACG 784
     ::   |||   :::       :::||||||   |||       :::::::|||
 444 alThrProGlnGlnProValThrLysAlaAlaThrThrThrAsnAlaThr 460

 785 GCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTG 834
                    ||||||:::         :::|||:::|||
 461 Pro.............ProProGlnProIleProLysAlaAlaThrThrTh 473

 835 GCAGATATGGCGCAACTCAAAGACTATGCCGCAGCAG.............. 871
                     :::   :::   |||||||||
 473 rThrAla.........ThrProValThrProGlnGlnProIleProLysA 487

 871 ................................................. 871

 487 laGlyThrAspAlaAlaProProProAlaValProLysAlaProSerAsp 503

 872 ........................CCATCCGCGATTGGGCAGTCCAAAAC 897
                             |||:::|||      :::   :::
 504 GlyArgAlaAlaThrProGlyValProAsnAlaAlaThrAspProGlnLy 520

 898 CCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGC 947
      |||   |||                |||||||||::::::      :::|
 520 sProProProThrProGlnSerValProSerAlaValThrGluProLysP 537

 948 CATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGGCG 997
     ||:::|||
 537 roGlnPro.......................................... 539

 998 GCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTG 1047

 539 ................................................. 539

1048 CCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGCATACGC 1097
      |||   :::   |||||||||            |||         :::|
 540 .ArgAlaAlaProProProSerAsnGluAlaThrProAlaValProSerP 556

1098 CAAATACCCGT.........CCCCTT...ACCATTCCCGAA......... 1126
     ||:::   :::         ||||||   |||||||||:::
 556 roSerProAsnLeuLysSerProLeuProThrIleProLysProValPro 572

1127 ........................ATATCCGTTCAAACTTGGAGCAGCGT 1152
                              :::::::::|||      ::::::::
 573 LeuMetAlaLeuThrProGlnProValThrAlaGlnMetValThrGlnLe 589

1153 TACGGCAAAGAAAACATCACCTCCTCAACCGTGCCGCCGTCAAACGGCAA 1202
     ::::|||   |||   ||||||      |||                |||
 589 uAlaAlaThrLysProSerProIleValProLysAlaSerProLysAlaL 606

1203 AAATGTCA 1210
        |||:::
 606 euMetThr 608
```

seq_name: pir2:H87520

seq_documentation_block:
hypothetical protein CC2193 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: H87520
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kon, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: H87520
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-424 <STO>
A;Cross-references: GB:AE005673; NID:g13423694; PIDN:AAK24164.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2193

alignment_scores:
Quality: 133.00 Length: 489
Ratio: 0.627 Gaps: 23
Percent Similarity: 43.354 Percent Identity: 21.677

alignment_block:
US-09-303-518D-463 x H87520 ..

Align seg 1/1 to: H87520 from: 1 to: 424

```
 303 CCATTCGCCCTTCGACAACCATGCCTCACATTCCGATTCTGACGAAGCCG 352
     ||||||:::   |||:::|||
   5 ProPheValSerArgArgPro............................ 11

 353 GTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCCATTGGGACGGATAC 402
                                                 |||   |
  12 .....................................AspArgArgA 15

 403 GAACACCATCCCGCCGACGGCTATGACGGGCCACAGGGCGG......... 443
     ||:::|||:::||||||         |||      |||
  15 rgAlaProAlaArgArgAlaArgAlaArgArgProGlyProPheAlaLeu 31

 444 ..................CGGCTATCCCGCTCCCAAAGGCGCGAGGGA.. 473
                         |||      |||   |||   |||:::|||
  32 SerValAlaAspArgValArgCysValArgProGlnGlyArgHisGlyAl 48

 474 ..........TATATACAGCTACGACATAAAAGG.........CGTTGCC 504
             :::   :::|||   |||   |||         :::
  48 aGlyGlyLeuHisArgGluLeuLeuHisAlaArgGlyGlyAspGluIleA 65

 505 CAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAACGGCT 554
       :::         :::       |||         ||||||      :::
  65 rgArgLeuArgArgAspArgAlaArgTrpArgProHisArgProGlySer 81

 555 TGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTAGGCG 604
                         |||         :::|||   |||:::   ||
  82 GlyLeuHisAlaArgAlaCysAlaGlyGlyGlyAspAspArgGluAlaAr 98

 605 ACGGATTCAAACGCGCCAC......CCGATACAGCCCCGAGCT...GGAC 645
     |:::::::::   :::|||         :::|||||||::   |||
  98 gHisLeuSerAspHisHisArgSerAlaArgArgProArgGlyGlnGlyG 115
```

```
 646 AGATCGGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAA 695
     :::|||||   |||   ::::::   |||:::   :::
 115 lyLeuGlyGlnGlyArgValGlyValLeuArgArgLeuGlnGlyAlaGly 131

 696 AAACATCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGC 745
     ::::::|||||:::::::::   :::      ||||||   |||   ||
 132 ArgArgHisArgHisAlaGluProAspProGlyArgArgLeuCysAspAl 148

 746 AGGGTATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTG 795
     |         ||||||   ::::::|||           ||||||:::|
 148 a.........ArgArgAlaGlnArgCys.........GlyLeuGlyAlaA 159

 796 CTTTCCACCGAAAACAAGATGGCGCGCATCAACGATTTGGCAGATATGGC 845
     ||      |||:::   :::|||:::            ||||||
 159 la......ArgGlnGlySerGlyValPro.........GlyArgAspPro 170

 846 GCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAA 895
     :::   |||         |||   ::::::         |||:::   ::
 171 SerGlnGlnAspArgHisArgProGlyArgArgValArgGlyAlaValAr 187

 896 ACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCA 945
     :   |||:::|||   |||                          ::::
 187 gArgGlnAlaArgGlyArgPro.......................AlaA 196

 946 GCCATCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAATACGGCTTGGG 995
     ::||||||   :::|||                  :::      |||
 196 snHisProAlaArgArg.................ArgArgProLeuArg 206

 996 CGGCATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCAT 1045
     ||||||:::|||   :::   :::|||
 207 ArgHisGlnGlyGlyAlaGlnAspAla..................... 215

1046 TGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTGCCGATGCGGC.... 1091
              |||   ||||||||||||   :::   |||   |||
 216 ..........GluGluArgArgGlnArgProLeuArgArgProGlyArgA 229

1092 ..ATACGCCAAATACCCGTCCCCTTACCATTCCCGAAA.........TAT 1130
       :::   :::::::|||   |||:::   :::|||:::
 229 laValGlyArgLeuProAspProValAlaLeuProArgArgArgAlaArg 245

1131 CCGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAA 1180
     |||         |||      |||:::   |||           |||::
 246 ProArgProAlaGlyProGluLeuHisLeuArgProAsp......LeuAr 260

1181 CCGTGCCGCCG..................................... 1191
     : :::|||||||
 260 g.IleProProLeuSerGluArgProValMetLysThrArgThrLeuLeu 276

1192 .............................................TC 1193
                                                   ||
 277 ThrAlaAlaValAlaLeuSerLeuIleAlaGlyAlaAlaHisAlaAlaSe 293

1194 AAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACCCGAAGACAGGCG 1243
     |:::::::::::::::   :::   :::|||   :::
 293 rHisAlaArgAspIlePheIleLysGluGlnAspGln............. 305

1244 TACCGTTTGACGGTAAAGGGTTTCCGAATTTTGAGAAGCAC......... 1284
            :::|||   ||||||   :::   :::::::::
 306 ........AsnGlyAspGlyPheValThrLysAspGluTyrAlaAlaThr 319

1285 ..............GTGAAATATGATACGAAGCTCGATATTCAAGAATT 1319
                    :::|||   ||||||
 320 ArgAlaIleGlnPheAlaLysThrAspThrAsp................. 330

1320 ATCGGGGGGCGGTATACCTAAGGCTAAGCCTGTGTTTGATGCGAAACCGA 1369
      ::::::|||:::::   :::|||:::   |||   :::   |||   |
 331 .AsnSerGlyAlaLeuSerGlnAlaGluTyrValAlaGluPheLysAlaA 347

1370 GATGGGAGGTTGATAGGAAGCTTAATAAATTGACAACTCGTGAGCAGGTG 1419
     ||   |||:::          :::   :::|||   ::::::::::::
 347 rgLeuGluAlaLysLeuThrAlaSerAspMetThrProGluLysLysAla 363

1420 GAGAAAAATGTTCAGGAAACGAGAAGAAGG..................... 1449
```

```
          |||:::::::::::::::   |||:::
      364 GluGluArgMetArgGlnMetArgGlnThrAspValArgPheGlyValLe 380

     1450 ....AGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGGGAAA 1495
              |||::::::|||         :::   |||:::   ::::::::::
      380 uAspSerAspLysSerGlyGlyIleThrLysAlaGluPheAspTyrSerG 397

     1496 ATAAAACAGGGTTAGATTTTAATCATTTTATAGGTGGTGATATCAATAAG 1545
          ::                 ::::::   |||:::      |||   |||
      397 ly..............TrpArgMetPheValThrHisAspThrAsnThr 408

     1546 AAAGGCACAGTAACA 1560
             |||:::|||:::
      409 AspGlyAlaValSer 413
```

seq_name: pir2:A29130

seq_documentation_block:
beta-amylase (EC 3.2.1.2) / alpha-amylase (EC 3.2.1.1) precursor - Bacillus polymyxa
C;Species: Bacillus polymyxa
C;Date: 25-Oct-1987 #sequence_revision 03-Mar-1994 #text_change 15-Oct-1999
C;Accession: A29130; B29130; A32251; A29108
R;Kawazu, T.; Nakanishi, Y.; Uozumi, N.; Sasaki, T.; Yamagata, H.; Tsukagoshi, N.; Ud
J. Bacteriol. 169, 1564-1570, 1987
A;Title: Cloning and nucleotide sequence of the gene coding for enzymatically active
A;Reference number: A29130; MUID:87165765
A;Accession: A29130
A;Molecule type: DNA
A;Residues: 1-936 <KAW>
A;Cross-references: GB:M15817
A;Experimental source: strain 72
A;Accession: B29130
A;Molecule type: protein
A;Residues: 36-50 <KA2>
R;Uozumi, N.; Sakurai, K.; Sasaki, T.; Takekawa, S.; Yamagata, H.; Tsukagoshi, N.; Ud
J. Bacteriol. 171, 375-382, 1989
A;Title: A single gene directs synthesis of a precursor protein with beta- and alpha-
A;Reference number: A32251; MUID:89123046
A;Accession: A32251
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 937-1196 <UOZ>
A;Experimental source: strain 72
R;Rhodes, C.; Strasser, J.; Friedberg, F.
Nucleic Acids Res. 15, 3934, 1987
A;Title: Sequence of an active fragment of B. polymyxa beta amylase.
A;Reference number: A29108; MUID:87231094
A;Accession: A29108
A;Molecule type: DNA
A;Residues: 'MIGL',2-66,'S',68-99,'D',101-153,'N',155-176,'Q',178-226,'KS',229-329,'S
,'K',737-740,'S',742-757,'N',759-776 <RHO>
A;Experimental source: ATCC 8523
C;Genetics:
A;Start codon: TTG
C;Function:
A;Description: catalyzes both the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
vely from the non-reducing end (beta-amylase activity)
A;Pathway: glycogen/starch degradation
C;Superfamily: alpha-amylase core homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;1-35/Domain: signal sequence #status predicted <SIG>
F;36-1196/Product: beta-amylase/alpha-amylase #status predicted <MAT>
F;908-1033/Domain: alpha-amylase core homology <AMY>

alignment_scores:
            Quality: 133.00   Length: 717
              Ratio: 0.426   Gaps: 38
 Percent Similarity: 43.515   Percent Identity: 21.060

alignment_block:
 US-09-303-518D-463 x A29130   ..
```
```

Align seg 1/1 to: A29130 from: 1 to: 1196

```
 71 CCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGACCGTCAG 120
    |||:::|||:::           |||||||||
187 ProLysIleTyr...........LeuSerGly................. 193

121 CATTTCGAACCCGACGGGAAAT...........ACCACCTATTCGGCAG 158
        :::|||:::|||:::          :::   |||   |||| \
194 .......GlyProSerGlyGluLeuArgTyrProSerTyrTyrProAlaA 208

159 CAGGGGGGAGCTTGCCNAGC............................. 178
    |||||   |||   |||:::
208 laGlyTrpSerTyrProGlyArgGlyLysPheGlnAlaTyrThrGluThr 224

179 ..............GCAACGGCCATATCGGAT................. 196
                   ||||||::::::|||
225 AlaLysAsnAlaPheArgThrAlaMetAsnAspLysTyrGlySerLeuAs 241

197 ................TGGGAAACATACAAAGCCATCA.........GTT 221
                  |||   |||          |||         :::
241 pLysIleAsnAlaAlaTrpGlyThr.LysLeuThrSerLeuSerGlnIle 257

222 GGGCCACCTGATGAT..............TCAACAGGCGGCCGTTGAAG 256
    :::||||||   |||             ::::::||||||
258 AsnProProThrAspGlyAspGlyPheTyrThrAsnGlyGlyTyrAsnSe 274

257 GAAATATCGGCTACATTGTCCGCTTTTCCGATCACGGGCACAAATTCCAT 306
        |||  |||              :::|||           ::::::
274 rAlaTyr.Gly..............LysAspPheLeuSerTrpTyrGln 285

307 TCGCCCTTCGACAACCAT.................GCCTCACATTCCGA 338
    |||   :::::::::|||                 |||:::|||:::::
286 SerValLeuGluLysHisLeuGlyValIleGlyAlaAlaAlaHisLysAs 302

339 TTCTGACGAAGCCGGTAGTCCCGTTGACGGATTCAGCCTTTACCGCATCC 388
    :   |||::::::   :::         |||   ::::::      :::|
302 nPheAspSerValPheGlyValArgIleGlyAlaLysIleSerGlyLeuH 319

389 ATTGGGACGGATACGAACACCATCCCGCCGACGGCTATGACGGGCCACAG 438
    ||||||:::          ::::::||||||      :::         |||
319 isTrpGln......MetAsnAsnProAlaMetProHisGlyThrGluGln 333

439 GGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGATATATACAGCTACGA 488
    :::|||||||||                                |||:::|||::
334 AlaGlyGlyTyr.........................TyrAspTyrAs 341

489 CATAAAAGGCGTTGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCA 538
    :
341 n.............................................. 341

539 GCACCGGACAACGGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTG 588
              ||||||   ::::::|||   :::|||         |||
342 ..........ArgLeuIleGlnLysPheLysAspAlaAspLeuAspLeu 354

589 ACGCAAGGAGTAGGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGA 638
                                        |||          ||
355 ............................ThrPheThrCysLeuGl 360

639 GCTGGACAGATCGGGCAATGCCGCCGAAGCCTTC......AACGGCACTG 682
    |::::::   ||||||:::|||   :::             :::      :
360 uMetSerAspSerGlyThrAlaProAsnTyrSerLeuProSerThrLeuV 377

683 CAGATATCGTCAAAAACATCATCGGCGCGGCAGGA.............. 717
    ::|||   |||:::::::|||   :::|||   |||
377 alAspThrValSerSerIleAlaAsnAlaLysGlyValArgLeuAsnGly 393

718 ......GAAATTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGG 761
          :::      :::|||::::::   |||   |||:::|||
394 GluAsnAlaLeuProThrGlyGlySerGlyPheGlnLysIleGluGluLy 410

762 CTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACA 811
       :::           ||||||:::   ||||||
410 sIleThrLysPheGlyTyrHisGlyPheThrLeuLeu............. 422

812 AGATGGCGCGCATCAACGATTTG.....................GCA 837
           |||||||||:::|||                       :::
423 ........ArgIleAsnAsnLeuValAsnAsnAspGlySerProThrGly 436

838 GATATGGCGCAACTCAAAGACTATGCCGCAGCAGCCATCCGCGATTGGGC 887
    :::::::::   :::|||:::|||       :::       :::
437 GluLeuSerGlyPheLysGlnTyrIleIleSerLysAlaLys........ 450

888 AGTCCAAAACCCCAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCT 937
              |||:::        |||                :::|||:
451 ..........ProAspAsnAsnGlyGlyThrGlyAsnLysValThrIleT 464

938 TTATGGCAGCCATCCCCATCAAAGGGATTGGAGCT......GTCCGGGGA 981
    ::                  ||||||:::::::::      ::::::
464 yrTyrLys............LysGlyPheAsnSerProTyrIleHisTyr 476

982 AAATACGGCTTGGGCGGCATCACGGCACATCCTGTCAAGCGGTCGCAGAT 1031
    :::   :::   |||:::   ||||||   |||       :::   |||
477 ArgProAlaGlyGlySerTrpThrAlaAlaProGlyValLysMetGlnAs 493

1032 GGGCGCGATCGCATTGCCGAAAGGGAAATCCGCCGTCAGCGACAATTTTG 1081
     :::   |||:::         |||   :::   ::::::   :::::::
493  pAlaGluIleSer.........GlyTyrAlaLysIleThrValAspIleG 507

1082 CCGATGCGGCATACGCCAAATACCCGTCCCCTTACCATTCCCGAAATATC 1131
     :::::|||:::      :::                    ::::::|||
507  lySerAlaSerGlnLeuGluAlaAlaPheAsnAspGlyAsnAsnAsnTrp 523

1132 CGTTCAAACTTGGAGCAGCGTTACGGCAAAGAAAACATCACCTCCTCAAC 1181
        ||||||      ::::::|||:::   ::::::   |||   |||||
524  AspSerAsnAsnThrLysAsnTyrSerPheSerThrGlyThr...SerTh 539

1182 CGTGCCGCCGTCAAACGGCAAAAATGTCAAACTGGCAGACCAACGCCACC 1231
     |      |||:::|||:::   |||:::          :::       |
539  rTyrThrProGlyAsnSerGlyAsnAlaGlyThrIleThrSerGlyAlaP 556

1232 CGAAGACAGGCGTACCGTTTGACGGTAAAGGGTTTCCGAAT......... 1272
     ||      :::   |||   ||||||   |||      |||
556  roAlaGlyAlaAsnProGlyAspGlyGlyGlyThrThrAsnLysValThr 572

1273 ......TTTGAGAAG..............CACGTGAAATAT........ 1293
           ::::::|||               ::::::   |||
573  ValTyrTyrLysLysGlyPheAsnSerProTyrIleHisTyrArgProAl 589

1294 ........................GATACGAAGCTCGATATTCAAGAAT 1318
                              :::|||::::::      |||:
589  aGlyGlySerTrpThrAlaAlaProGlyValLysMetGlnAspAlaGluI 606

1319 TATCGGGG.....................GGCGGTATACCTAAGGCT 1344
     ::||||||                    |||:::      :::
606  leSerGlyTyrAlaLysIleThrValAspIleGlySerAlaSerGlnLeu 622

1345 AAGCCTGTGTTT...GATGCGAAACCGAGATGGGAGGTTGATAGGAAGCT 1391
     :::   :::|||   |||::::::   :::|||:::   ::::::
623  GluAlaAlaPheAsnAspGlyAsnAsnAsnTrpAspSerAsnAsnThrLy 639

1392 TAATAAATTGACAACTCGTGAGCAGGTGGAGAAAAATGTTCAGGAAACGA 1441
      |||   |||   :::                          |||
639  sAsnTyrLeuPheSer.....................ThrGlyThrS 648

1442 GAAGAAGGAGTCAGAGTAGTCAGTTTAAAGCCCATGCGCAACGAGAATGG 1491
           :::   :::|||:::      |||   :::   |||
648  erThrTyrThrProGlySerAsnGlyAlaAlaGlyThrIleArg...... 662

1492 GAAAATAAAACAGGGTTAGATTTTAATCATTTTATA.............. 1527
              ||||||         :::      :::
663  .........ThrGlyAlaProSerGlySerValLeuSerValValThrSe 676

1528 .......GGTGGTGATATCAATAAGAAAGGCACAGTAACAGGA....... 1563
           :::   |||:::|||:::         |||||||||
676  rThrTyrAlaThrAspLeuAsnGlu.........ValThrGlyProIleG 690
```

```
1564 ..GGGCATAGTCTAACCCGTGGTGATGTACGGGTGATACAACAAACCTCG 1611
              ::::::|||:::      :::::::::|||      :::|||
 690 lnThrGluLysLeuSerGlyValSerLeuAsnValSerThrSerThrTyr 706

1612 GCACCTGATAAACATGGGGTTTATCAAGCGACAGTGGAAATTAAAAAGCC 1661
     ||||||:::::::::||||||   ::::::|||::::::   :::   ||
 707 AlaProAsnSerAsnGlyVal...GluValThrAlaGlnThrGluAlaPr 722

1662 TGATGGAAGTTGGGAGGTGAAAACGAAAAAAGGTGGGAAAGTGATGACCA 1711
     |:::|||::::::         |||:::            ::::::::::
 722 oSerGlyAlaPhe.........ThrSerMetAspLeuGlyThrLeuSerA 736

1712 AGCACACCATGTTCCCAAAAGATTGGGATGAGGCTAGAATTAGGGCTGAA 1761
     ::   |||   :::      ||||||::::::      |||
 736 snProThrSerLeuAsnThrAspTrpSerLysGlnSerIleTyrPheIle 752

1762 GTTACTTCGGCTTGGGAAAGTAGAATAATGCTTAAGGATAAT........ 1803
     :::|||:::   :::::::::            :::||||||
 753 MetThrAspArgPheSerAsnGlyAspProSerAsnAspAsnTyrGlyGl 769

1804 ..........................AAATGGCAGGGTACAAGTAAAT 1825
                               ||||||:::|||   :::   :
 769 yPheAsnSerAsnAsnSerAspGlnArgLysTrpHisGlyGlyAspPheG 786

1826 CGGGT.....................ATTAAAATAGAAGGATTTACC 1851
     ::|||                    ||||||      |||||||||
 786 lnGlyIleIleAsnLysLeuAspTyrIleLysAsnMetGlyPheThr 801
```

seq_name: pir2:G01763

seq_documentation_block:
atrophin-1 - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C;Accession: G01763
R;Margolis, R.L.
submitted to the EMBL Data Library, March 1995
A;Reference number: G08343
A;Accession: G01763
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1184 <MAR>
A;Cross-references: EMBL:U23851; NID:g915325; PID:g915326
C;Genetics:
A;Gene: GDB:DRPLA; B37
A;Cross-references: GDB:270336; OMIM:125370
A;Map position: 12p-12p

alignment_scores:
Quality: 132.00 Length: 458
Ratio: 0.714 Gaps: 22
Percent Similarity: 40.393 Percent Identity: 22.052

alignment_block:
US-09-303-518D-463 x G01763 ..

Align seg 1/1 to: G01763 from: 1 to: 1184

```
  65 CACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGAC 114
     |||   |||           |||   |||:::|||       |||
 252 HisProPro.........ProThrThrProIleSerValSerSerSerGl 265

 115 CGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGG 164
      :::|||   :::   ||||||        |||      ::::::||||
 265 yAlaSerGlyAlaProProThrLysProProThrThrProValGlyGlyG 282

 165 GGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAAGCC 214
     ||:::|.||||||||||         :::::::::    |||:::   ||
 282 lyAsnLeuProSerAlaProProProAlaAsnPhe.ProHisValThrPr 298

 215 ATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATC 264
     |::::::   ||||||      :::     |||||||||:::::   ||
 298 oAsnLeu...ProProProProAlaLeu..ArgProLeuAsnAsnAlaSe 313

 265 GGCTACA.........TTGTCCGCTTTTCCGATCACGGGC.......... 295
     ||||:::         |||:::|||   |||:::   |||
 313 rAlaSerProProGlyLeuGlyAlaGlnProLeuProGlyHisLeuProS 330

 295 ..................................................... 295

 330 erProHisAlaMetGlyGlnGlyIleGlyGlyLeuProProGlyProGlu 346

 296 ......................ACAAATTCCATTCGCCCTTCGACAAC 321
                                :::|||:::   |||:::::::::
 347 LysGlyProThrLeuAlaProSerProHisSerLeuProProAlaSerSe 363

 322 CATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATT 371
     :           |||   :::      |||            ::::::|
 363 rSerAlaProAlaProProMetArgPheProTyrSerSerSerSerSerS 380

 372 CAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACG 421
     ||:::   :::||||||   :::::::::::::::::
 380 erSerAlaAlaAlaSerSerSerSerSerSerSerSer............ 392

 422 GCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGG 471
                       ::::::|||   |||:::|||   :::
 393 .................SerSerAlaSerProPheProAlaSer..... 401

 472 GATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAA 521
                           :::|||||||||:::         |||
 402 .....................GlnAlaLeuProSerTyrProHisSerP 411

 522 CCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACA 571
           |||   |||:::         ::::::   |||         |||
 411 heProProProThrSerLeuSerValSerAsnGlnProProLysTyrThr 427

 572 ATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCC 621
     :::|||:::|||
 428 GlnProSerLeu..................................... 431

 622 ACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTT 671
      |||:::   |||      |||::::::         ||||||   |||
 432 .ProSerGlnAla...ValTrpSerGlnGlyProProProProProProT 447

 672 CAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAA 721
             |||      ::|||
 447 yrGlyArgLeuLeuAlaAsnSer........................... 454

 722 TTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATT 771
        :::|||:::      |||            :::   ||||||:::
 455 ...AsnAlaHisProGlyProProProProSerThrGlyAlaGlnSer.T 470

 772 GCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCG 821
       :::|||      :::::::::     |||:::  |||:::
 470 hrAlaHisProProValSerThrHisHisHisHisGlnGlnGlnGln 486

 822 CATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAGCAG 871
     :::|||:::                    |||:::      ::::::
 487 GlnGlnGln.................GlnGlnGlnGlnGlnGlnGlnHi 497

 872 CCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAA 921
      |||   :::   |||   |||   |||
 497 sHisGlyAsnSerGlyProProProPro..................... 506

 922 GCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGC 971
                        |||:::   ||||||      :::   :::
 507 ...................GlyAlaPheProHisProLeuGluGlyGly 516

 972 TGTCCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCT....... 1014
                           :::   ||||||  |||||||||
 517 ...................SerSerHisHis.AlaHisProTyrAlaM 526

1014 ..................................................... 1014

 526 etSerProSerLeuGlySerLeuArgProTyrProProGlyProAlaHis 542
```

```
 1015 .....................GTCAAGCGGTCGCAGATGGGCGCGATCGC 1043
                           |||:::  ||||||  |||     ::
  543 LeuProProProHisSerGlnValSerTyrSerGlnAlaGlyProAsnGl 559

 1044 ATTGCCGAAAGGGAAATCCGCC...GTCAGCGACAATTTTGCCGATGCGG 1090
      :  |||  ::::::|||:::     |||::::::  ::::::::::
  559 yProProValSerSerSerSerAsnSerSerSerSerThrSerGlnGlyS 576

 1091 CATAC.....................................GCCAAA 1101
      ::|||                                     |||
  576 erTyrProCysSerHisProSerProSerGlnGlyProGlnGlyAlaPro 592

 1102 TACCCGTCCCCTTACCATTCCCGAAATA......TCCGTTCAAACTTGGA 1145
      ||||||  |||   :::|||  :::     |||     |||  |
  593 TyrProPhePro.ProValProThrValThrThrSerSerAlaThrLeuS 609

 1146 GCAGCGTTACGGCA.........................AAGAAAACA 1168
      ||:::|||  |||                         |||  :::
  609 erThrValIleAlaThrValAlaSerSerProAlaGlyTyrLysThrAla 625

 1169 TCACCTCCTCAACCG 1183
      |||||||||  |||
  626 SerProProGlyPro 630

seq_name: pir2:S50832

seq_documentation_block:
atrophin-1 - human
C;Species: Homo sapiens (man)
C;Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 07-May-1999
C;Accession: S50832
R;Nagafuchi, S.; Yanagisawa, H.; Ohsaki, E.; Shirayama, T.; Tadokoro, K.; Inoue, T.; Yam
Nature Genet. 8, 177-181, 1994
A;Title: Structure and expression of the gene responsible for the triplet repeat disorde
A;Reference number: S50832; MUID:95144175
A;Accession: S50832
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1184 <NAG>
A;Cross-references: EMBL:D31840
C;Genetics:
A;Gene: GDB:DRPLA; B37
A;Cross-references: GDB:270336; OMIM:125370
A;Map position: 12p13.31-12p13.3112p-12p


alignment_scores:
            Quality:  132.00            Length:    458
              Ratio:  0.714              Gaps:     22
 Percent Similarity:  40.393   Percent Identity: 22.052

alignment_block:
 US-09-303-518D-463 x S50832   ..

 Align seg 1/1  to: S50832  from: 1  to: 1184

   65 CACACGCCTCAGATTTGGCAAACGATCCCTTTATCCGGCAGGTTCTCGAC 114
      |||  |||           |||  |||:::|||        |||
  253 HisProPro.........ProThrThrProIleSerValSerSerSerGl 266

  115 CGTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGG 164
      :::|||  :::  ||||||        |||     ::::::||||
  266 yAlaSerGlyAlaProProThrLysProProThrThrProValGlyGlyG 283

  165 GGAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAAGCC 214
      ||:::||||||||||||        :::::::::   |||:::  ||
  283 lyAsnLeuProSerAlaProProProAlaAsnPhe.ProHisValThrPr 299

  215 ATCAGTTGGGCCACCTGATGATTCAACAGGCGGCCGTTGAAGGAAATATC 264
      |::::::  ||||||     :::     |||||||||::::::  ||
  299 oAsnLeu...ProProProProAlaLeu..ArgProLeuAsnAsnAlaSe 314

  265 GGCTACA.........TTGTCCGCTTTTCCGATCACGGGC.......... 295
      ||||:::         |||:::|||  |||:::  |||
  314 rAlaSerProProGlyLeuGlyAlaGlnProLeuProGlyHisLeuProS 331

  295 ..................................................... 295

  331 erProTyrAlaMetGlyGlnGlyMetGlyGlyLeuProProGlyProGlu 347

  296 ........................ACAAATTCCATTCGCCCTTCGACAAC 321
                               :::|||:::  |||::::::::
  348 LysGlyProThrLeuAlaProSerProHisSerLeuProProAlaSerSe 364

  322 CATGCCTCACATTCCGATTCTGACGAAGCCGGTAGTCCCGTTGACGGATT 371
      :           |||  :::      |||           ::::::|
  364 rSerAlaProAlaProProMetArgPheProTyrSerSerSerSerSerS 381

  372 CAGCCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACG 421
      ||:::  :::||||||  ::::::::::::::::::
  381 erSerAlaAlaAlaSerSerSerSerSerSerSerSer............ 393

  422 GCTATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGG 471
                      ::::::|||  |||:::|||  :::
  394 ..................SerSerAlaSerProPheProAlaSer..... 402

  472 GATATATACAGCTACGACATAAAAGGCGTTGCCCAAAATATCCGCCTCAA 521
                               :::||||||||:::         |||
  403 .....................GlnAlaLeuProSerTyrProHisSerP 412

  522 CCTGACCGACAACCGCAGCACCGGACAACGGCTTGCCGACCGTTTCCACA 571
          |||  |||:::         ::::::  |||         |||
  412 heProProProThrSerLeuSerValSerAsnGlnProProLysTyrThr 428

  572 ATGCCGGCGCTATGCTGACGCAAGGAGTAGGCGACGGATTCAAACGCGCC 621
      :::|||:::|||
  429 GlnProSerLeu...................................... 432

  622 ACCCGATACAGCCCCGAGCTGGACAGATCGGGCAATGCCGCCGAAGCCTT 671
       |||:::  |||      |||::::::         ||||||  |||
  433 .ProSerGlnAla...ValTrpSerGlnGlyProProProProProProT 448

  672 CAACGGCACTGCAGATATCGTCAAAAACATCATCGGCGCGGCAGGAGAAA 721
            |||      :::|||
  448 yrGlyArgLeuLeuAlaAsnSer.......................... 455

  722 TTGTCGGCGCAGGCGATGCCGTGCAGGGTATAAGCGAAGGCTCAAACATT 771
      :::|||:::      |||           :::  ||||||:::
  456 ...AsnAlaHisProGlyProProheProProSerThrGlyAlaGlnSer.T 471

  772 GCTGTCATGCACGGCTTGGGTCTGCTTTCCACCGAAAACAAGATGGCGCG 821
      :::|||      :::::::::      |||:::  |||:::
  471 hrAlaHisProProValSerThrHisHisHisHisGlnGlnGlnGlnGln 487

  822 CATCAACGATTTGGCAGATATGGCGCAACTCAAAGACTATGCCGCAGCAG 871
      :::|||:::                      |||:::      ::::::
  488 GlnGlnGln.................GlnGlnGlnGlnGlnGlnGlnHi 498

  872 CCATCCGCGATTGGGCAGTCCAAAACCCCAATGCCGCACAAGGCATAGAA 921
       |||  :::  |||  |||  |||
  498 sHisGlyAsnSerGlyProProProPro..................... 507

  922 GCCGTCAGCAATATCTTTATGGCAGCCATCCCCATCAAAGGGATTGGAGC 971
                           |||:::  ||||||      :::  :::
  508 ..................GlyAlaPheProHisProLeuGluGlyGly 517

  972 TGTCCGGGGAAAATACGGCTTGGGCGGCATCACGGCACATCCT....... 1014
                        :::  ||||||  ||||||||
  518 ..................SerSerHisHis.AlaHisProTyrAlaM 527

 1014 ..................................................... 1014

  527 etSerProSerLeuGlySerLeuArgProTyrProProGlyProAlaHis 543

 1015 .....................GTCAAGCGGTCGCAGATGGGCGCGATCGC 1043
```

```
                              |||:::  ||||||  |||      ::
     544 LeuProProProHisSerGlnValSerTyrSerGlnAlaGlyProAsnGl 560

    1044 ATTGCCGAAAGGGAAATCCGCC...GTCAGCGACAATTTTGCCGATGCGG 1090
         :  |||  ::::::|||:::    |||::::::  ::::::::::
     560 yProProValSerSerSerSerAsnSerSerSerSerThrSerGlnGlyS 577

    1091 CATAC.....GCCAAA 1101
         ::|||          |||
     577 erTyrProCysSerHisProSerProSerGlnGlyProGlnGlyAlaPro 593

    1102 TACCCGTCCCCTTACCATTCCCGAAATA......TCCGTTCAAACTTGGA 1145
         ||||||  |||  :::|||  :::     |||     |||   |
     594 TyrProPhePro.ProValProThrValThrThrSerSerAlaThrLeuS 610

    1146 GCAGCGTTACGGCA.....AAGAAAACA 1168
         ||:::|||  |||              |||  :::
     610 erThrValIleAlaThrValAlaSerSerProAlaGlyTyrLysThrAla 626

    1169 TCACCTCCTCAACCG 1183
         |||||||||  |||
     627 SerProProGlyPro 631

seq_name: pir2:B84331

seq_documentation_block:
hypothetical protein Vng1794c [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: B84331
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
 Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483
A;Accession: B84331
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-694 <STO>
A;Cross-references: GB:AE004437; NID:g10581248; PIDN:AAG20014.1; GSPDB:GN00138
C;Genetics:
A;Gene: VNG1794C


alignment_scores:
           Quality:  131.50            Length:    410
             Ratio:  0.739              Gaps:     23
 Percent Similarity:  43.415   Percent Identity: 24.146

alignment_block:
 US-09-303-518D-463 x B84331   ..

 Align seg 1/1  to: B84331  from: 1  to: 694

     116 GTCAGCATTTCGAACCCGACGGGAAATACCACCTATTCGGCAGCAGGGGG 165
         |||:::   ::::::||||||:::   ||||||   :::      |||::
      72 ValAlaAlaAlaGluProThrAla...ThrThrValAspArgAspGlyAs 87

     166 GAGCTTGCCNAGCGCAACGGCCATATCGGATTGGGAAACATACAAA.... 211
         ::::      :::::||||||      |||   ||||||
      87 nAsp......AspThrThrAlaTyrProAspGlyGluThrIleProLeuM 102

     212 ..........GCCATCAGTTGGGCCACCTGA 232
                    |||      :::||| |||||
     102 etAlaValAspGlySerValValAlaPheAlaAlaProPheAla.ProAs 118

     233 TGATTCAACAGGCGGCCGTTGAAGGAAATATCGGCTACATTGTCCGC... 279
         |||||||   :::|||       ||||||        :::::::::
     118 pAspSerAspAlaGlySer.TyrGlyAsnGluAspValLeuLeuAsnVal 134

     280 .........TTTTCCGATCACGGGCACAAATTC.....CATTC 308

                 :::::::|||   |||      :::            |||::
     135 PheAspAlaTyrAlaAspAlaGlyThrIleTyrTrpAspAspGlyHisAs 151

     309 GCCCTTCGACAACCATGCCTCACATTCC.....GATTCTGACG 346
         :   |||   ::::::|||||||||||||              :::|||
     151 pGlnPheTyrAspArgAlaSerHisSerAlaValAlaAlaArgAlaAspT 168

     347 AAGCCGGT..... 354
           ||||||
     168 hrAlaGlyHisThrValAlaAlaThrThrAspIleAlaAspAlaValThr 184

     355 .....AGTCCCGTTGACGGATTCAG 374
                                  ||||||:::|||   |||||
     185 ArgGluAspAlaAlaAlaAlaValValThrSerProAlaAspProPheSe 201

     375 CCTTTACCGCATCCATTGGGACGGATACGAACACCATCCCGCCGACGGCT 424
         |           :::   |||:::                  ||||||
     201 r......AlaAlaGluGlnAspAlaLeuAlaAlaPheAlaAlaAsp.... 214

     425 ATGACGGGCCACAGGGCGGCGGCTATCCCGCTCCCAAAGGCGCGAGGGAT 474
                       ||||||      :::      :::      |||
     215 .................GlyGlyLeuValValLeuMetSerGlnSerAsp 225

     475 ATATACAGCTACGAC.....ATAAAAGGCGT 500
               :::||||||                  |||::
     226 TyrLysAsnTyrAspAlaThrAlaSerLeuAsnAspLeuAlaAspGlyIl 242

     501 TGCCCAAAATATCCGCCTCAACCTGACCGACAACCGCAGCACCGGACAAC 550
         :      :::||||||:::|||   |||                   :
     242 eAspThrThrIleArgPheAsnAspThr.....G 252

     551 GGCTTGCCGACCGTTTCCACAATGCCGGCGCTATGCTGACGCAAGGAGTA 600
         :::::::::|||      ||||||||||||||
     252 lnValThrAspAspGlnHisAsnAlaGlyAla..... 262

     601 GGCGACGGATTCAAACGCGCCACCCGATACAGCCCCGAGCTGGACAGATC 650
               :::|||      :::|||         |||:::   |||   ::
     263 ......AlaPheAlaProThrThrAlaThrPheProSerAspAspTyrAl 277

     651 GGGCAATGCCGCCGAAGCCTTCAACGGCACTGCAGATATCGTCAAAAACA 700
         ::::      |||         :::|||||   |||::::::
     277 aSerLeuLeuAlaAlaArgThrAspGlyThrAspAspValThr..... 291

     701 TCATCGGCGCGGCAGGAGAAATTGTCGGCGCAGGCGATGCCGTGCAGGGT 750
                  :::::::::   :::   :::   |||:::|||:::
     292 ........IleThrAlaAspGluThrTyrThrValAspValValArg... 304

     751 ATAAGCGAAGGCTCAAACATTGCTGTCATGCACGGCTTGGGTCTGCTT.. 798
         :::::::::|||:::::::::   |||          |||
     305 ValThrAspGlyAspThrValAspValArgPheArgAspGlyThrThrAs 321

     799 .....TCCACCGAAA 808
                        :::|||   :
     321 pThrValArgIleValGlyIleAspThrProGluThrGlyGlnThrThrG 338

     809 ACAAGATGGCGCGCATCAACGATTTGGCAGATATGGCGCAACTCAAA... 855
         ::::::::|||:::   :::   :::::::|||   |||   ||||||
     338 luArgValAlaGluTyrGluGlyIleThrAspAlaAlaAlaLeuLysAla 354

     856 ......GACTATGCCGCAGCAGCCATCCGCGATTGGGCAGTCCAAAACCC 899
               |||   |||:::|||      ||||||
     355 ArgGlyAspAlaAlaThrAlaTyrAlaArgAsp..... 365

     900 CAATGCCGCACAAGGCATAGAAGCCGTCAGCAATATCTTTATGGCAGCCA 949
               |||   |||   ::::::|||:::      |||   |||:::
     366 .......AlaLeuGlyGlyGlnThrValThrLeuSerPheAspAlaGly. 379

     950 TCCCCATCAAAGGGATTGGAGCTGTCCGGGGAAAAATACGGCTTGGGCGGC 999
                       |||   :::||||||   ||||||
     380 .................GlyProAlaArgGlyAspTyrGly..... 387

    1000 ATCACGGCACATCCTGTCAAGCGGTCGCAGATGGGCGCGATCGCATTGCC 1049
                                        :::|||   ::::::|||||
```

```
 388 ......................ArgLeuLeuGlyTyrLeuThrLeuPr 396

1050 GAAAGGG............AAATCCGCCGTCAGCGACAATTTTGCC.... 1083
     |   |||            :::   ||||||:::|||::::::|||
 396 oAspGlyThrThrTyrAsnArgArgAlaValAlaAspGlyMetAlaArgV 413

1084 .....GATGCGGCATACGCCAAATAC 1104
           :::|||::::::::::|||
 413 alTyrAlaSerAlaPheSerGlnTyr 421
```